Psychological Bulletin

CONTENTS

Published Bimonthly by the
American Psychological Association

VOL. 53, NO. 4 JULY, 1956

The Psychological Bulletin contains evaluative reviews of research literature and articles on research methodology in psychology. This Journal *does not publish reports of original research or original theoretical articles.*

Manuscripts should be sent to Wayne Dennis, Department of Psychology, Brooklyn College, Brooklyn 10, New York. Manuscripts should be sent to this address even during the editor's absence on leave. Manuscripts will be acknowledged at this address and the carbon copies forwarded to the editor. However, beginning on August 1, 1955 and until August 1, 1956 *correspondence* should be addressed to the editor at the Department of Psychology, American University, Beirut, Lebanon.

Preparation of articles for publication. Authors are strongly advised to follow the general directions given in the "Publication Manual of the American Psychological Association" (*Psychological Bulletin*, 1952, 49 [No. 4, Part 2], 389–449). Special attention should be given to the section on the preparation of the references (pp. 432–440), since this is a particular source of difficulty in long reviews of research literature. *All copy must be double spaced, including the references.* All manuscripts should be submitted *in duplicate*. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

Reprints. Fifty free offprints are given to contributors of articles and notes. Authors of early publication articles receive no gratis offprints.

Communications—including subscriptions, orders of back issues, and changes of address—should be addressed to the American Psychological Association, 1333 Sixteenth Street N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 15th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: $8.00 (Foreign $8.50). Single copies, $1.50.

PUBLISHED BIMONTHLY BY:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Menasha, Wisconsin
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Entered as second class mail matter at the post office at Washington, D.C., under the act of March 3, 1879. Additional entry at the post office at Menasha, Wisconsin. Acceptance for mailing at special rate of postage under the provisions of Sec. 34-40 Par. (D) provided for in Section 538, act of February 28, 1925, authorized August 6, 1947. Printed in U.S.A.

Vol. 53, No. 4 July, 1956

Psychological Bulletin

THE STRUCTURE OF INTELLECT

J. P. GUILFORD
University of Southern California

It is the purpose of this report to describe a developing picture of the structure of human, adult intellect, as seen in terms of factors. Although the picture is incomplete, presenting it at this time seems desirable for two reasons. The picture now includes about forty different factors, most of which are generally unfamiliar. Many have only recently been demonstrated. Enough of the intellectual factors are known to suggest strongly the outlines of a system. The system has interesting theoretical implications, and, by reason of certain vacancies that appear, it points to still undiscovered factors, somewhat as the chemist's periodic table has served to indicate unknown elements.

As the writer has emphasized before (**10**, **13**), psychology and psychologists since Binet have taken a much too restricted view of human intelligence. We do not need to go into the reasons here. They can be summed up in a positive manner by saying that in attempting to fathom the nature of intellect more attention should be given to the human adult, particularly the superior human adult. It is to such specimens that we must go, if we are to investigate intellectual qualities and functions in their greatest scope and variety.

The advent of multiple-factor analysis has done something to broaden and enrich our conception of human intelligence, but factor theory and the results of factor analysis have had little effect upon the practices of measurement of intelligence. We do have a great variety of tests in such intelligence scales as the Binet and its revisions and in the Wechsler scales, to be sure. Too commonly, however, a single score is the only information utilized, and this single score is usually dominated by variance in only one or two factors. There is some indication of more general use of part scores, as in connection with the Wechsler tests, but each of these scores is usually factorially complex and its psychological meaning is largely unknown as well as ambiguous. The list of factors that is to be presented in this article should clearly demonstrate the very limited information that a single score can give concerning an individual, and on the other hand, the rich possibilities that those factors offer for more complete and more meaningful assessments of the intellects of persons.

Some seven years ago the writer initiated research aimed essentially at the study of adult, human intelligence, in a project on "aptitudes of high-level personnel."[1] In some re-

[1] Project 150-044, under Contract N6onr-23810, with the Office of Naval Research, monitored by the Personnel and Training Branch. Among those who have made the most significant contributions to the project are: Raymond M. Berger, Paul R. Christen-

spects this has been a continuation of wartime research in the AAF Aviation Psychology Research Program (**21**). The project was initiated with the conviction that the full scope of human intellect had not yet been explored, by factor-analysis methods or by any other methods. Thinking abilities, which have played important roles in some definitions of intelligence, seemed to have been neglected; particularly abilities having to do with productive thinking. Accordingly, four areas of thinking were selected for study, arbitrarily designated as reasoning, creativity, planning, and evaluation. While abilities belong to the context of individual differences, they also imply psychological functions of individuals. Hence it was thought that the findings would have much to offer toward an understanding of human thinking of various kinds, including problem solving.

Space does not permit describing in detail the research procedures; they have been described in the various technical reports from the aptitudes project (**14**, **15**). It should be pointed out, however, that the factor analyses were done in a research design that includes experimental features. Each investigation starts by hypothesizing that certain unitary abilities (psychological factors) exist and that they have certain properties. Psychological tests are then selected, adapted, and constructed for each hypothesized factor in a way that should lead to a "yes" or "no" answer from the analysis. The results should show that the factor hypothesized does or does not exist and it does or does not have the properties suggested. Thus, the kind of psychological test is an important independent variable, more or less under the control of the investigator. Certain other experimental variables are held relatively constant—the testing conditions and certain population features, such as sex, age, education, and motivation. The examinees have been men who were previously selected for military training leading to an officer's commission and they have been tested under ordinary military discipline.

In his survey of aptitude factors, published in 1951, French (**8**) listed, among others, 18 or 19 factors that can be classified as intellectual. Our investigations of thinking abilities have verified and helped to clarify many of these factors, besides introducing approximately as many new ones. Other recent investigations have also contributed new information regarding factors. The list presented here comes from all these sources.

Classes of Intellectual Factors

Inspection of the total list shows that the intellectual factors fall into two major groups—thinking and memory factors. The great majority of them can be regarded as thinking factors. Within this group, a threefold division appears—cognition (discovery) factors, production factors, and evaluation factors. The production group can be significantly subdivided into a class of convergent-thinking abilities and a class of divergent-thinking abilities.[2]

Cognition (Discovery) Factors

The cognition factors have to do with becoming aware of mental items

sen, Andrew L. Comrey, Russel F. Green, Alfred F. Hertzka, Norman W. Kettner, and Robert C. Wilson. I am particularly indebted to Christensen and Kettner for reading the preliminary draft of this paper, and to Philip R. Merrifield, also, for making suggestions.

[2] In the system of the intellectual factors to be described here, the reader will find some striking similarities to a system developed in-

or constructs of one kind or another. In the tests of these factors, something must be comprehended, recognized, or discovered by the examinee. They represent functions on the receiving side of behavior sequences.

The cognition abilities can be differentiated along the lines of two major principles. For some time we have been aware that thinking factors tend to pair off according to the material or content used in the tests. For each factor of a certain kind found in verbal tests there seemed to be a mate found in tests composed of figures or designs. We found, for example, a factor called *eduction of perceptual relations*, parallel with a factor called *eduction of conceptual relations*; a factor called *perceptual foresight*, parallel to one called *conceptual foresight;* and a factor of *perceptual classification*, parallel with one of *conceptual classification*. Only recently there has been increasing evidence for a third content category. Factors were found in tests whose contents are letters, or equivalent symbols, where neither perceived form or figure nor verbal meaning is the basis of operation. Factors based upon this type of material have been found, parallel to other factors where the test content is figural or verbal. Thus a third content category seems necessary.

A second major principle by which cognition factors may be differentiated psychologically depends upon the kind of thing discovered; whether it is a relation, a class, or a pattern, and so on. Thus, for each combination of content and thing discovered, we have a potential factor. The cognition factors can therefore be arranged in a matrix as shown in Table 1. The third and fourth rows seem to be complete at the present time.

dependently by Burt (**2**). The similarities are support for the idea that a system does exist.

There are vacancies in the other four rows. With each factor name are usually given two representative tests by name to help give the factor operational meaning.[3] A word or two will be said in addition regarding the less familiar tests.[4]

It should not be surprising to find the factor of *verbal comprehension*, the best known, and the dominant one in verbal-intelligence tests generally, in the first row of the cognition factors and in the conceptual column. The fact that the cognition factors sometimes come in threes leads us to look for parallel factors for the perceptual and structural columns. One candidate for the perceptual cell in this row would be the well-known factor of *perceptual speed*. This factor has to do with discriminations of small differences in form rather than in awareness of total figures, hence it does not quite fill the requirement of parallel properties with *verbal comprehension*. A better factor for this purpose is the one Thurstone (**28**) called "speed and strength of closure," called *figural closure* in Table 1. For this factor, awareness of perceived objects from limited cues is the key property. The limitation of cues is necessary to make the test sufficiently difficult for testing purposes.

There is no known factor that seems to belong in the second column of the first row of Table 1. In generalizing the class of three such factors, and in differentiation from other classes in Table 1, it is clear that those in the first row have to do with awareness of items, elements, or things. To denote this category Spearman's term "fundament" has been adopted.

[3] It should not be inferred that these are the only kinds of tests related to the factor.

[4] For more complete descriptions of the tests see particularly (**14**, **17**, **21**).

TABLE 1

Cognition (Discovery) Factors

Type of thing known or discovered	Type of content		
	Figural	Structural	Conceptual
Fundaments	*Figural closure* Street Gestalt Completion Mutilated Words		*Verbal comprehension* Vocabulary
Classes	*Perceptual classification* Figure Classification Picture Classification		*Verbal classification* Word Classification Verbal Classification
Relations	*Eduction of perceptual relations* Figure Analogies Figure Matrix	*Eduction of structural relations* Seeing Trends II Correlate Completion II	*Eduction of conceptual relations* Verbal Analogies Word Matrix
Patterns or systems	*Spatial orientation* Spatial Orientation Flags, Figures, Cards	*Eduction of patterns* Circle Reasoning Letter Triangle	*General reasoning* Arithmetic Reasoning Ship Destination
Problems			*Sensitivity to problems* Seeing Problems Seeing Deficiencies
Implications	*Perceptual foresight* Competitive Planning Route Planning		*Conceptual foresight* Pertinent Questions Alternate Methods *Penetration* Social Institutions Similarities

Two factors involving ability to recognize classes are known, one in which the class is formed on the basis of figural properties and the other on the basis of meanings. It was interesting that the Picture Classification test had more relation to the *perceptual-classification* factor than to the *conceptual-classification* factor in spite of the fact that the things to be classified were common objects, the basis for whose classification was intended to be their meanings. This might mean that the perceptual-conceptual distinction is a somewhat superficial matter, pertaining only to how the material is presented. It is possible, however, that in many of the items in this test the general shapes and sizes and other figural properties are an aid in classification. For example, there are cleaning implements, containers, etc., in some items, where similarities of appearance may serve as clues.

The difference between the Word Classification test and the Verbal Classification test is largely in the form of presentation of the problems. A sample item from the Word Classification test is: "A. horse B. cow C. man D. flower." Which word does not belong? In the Verbal Classification test, two short lists of words are

given to establish two classes, e.g., animals and pieces of furniture. A longer list of words is given, each one of which must be marked as belonging to one class or the other or to neither class.

Is there likely to be a factor having to do with the seeing of classes when class membership depends upon structural properties? Such a factor would be reasonable. We have much to learn regarding the scope of structural ideas. Thus far, structural factors have been found only in tests utilizing letters and very simple forms such as circles, dashes, and the like. One can raise the question whether mechanical conceptions, for example, belong in this class. There is also the question of where figural properties end and structural properties begin, also of where structural properties end and conceptual properties begin. We may actually have a continuum here. With respect to some categories (including classes, fundaments, etc.) there may be a rapid transition from figural to conceptual, thus leaving no basis for a third factor. It is likely that the factors in any row of Table **1** are positively and sometimes even substantially correlated. The general question of correlations among factors will be left for later discussion.

We have a complete triad of factors having to do with the seeing of relationships and tests to measure them that are similar except for content. The analogies tests are well known. A matrix test is essentially a two-dimensional analogies test, examples of which may be found in the Raven Progressive Matrices series. In the test Seeing Trends II, we find the following type of item: "anger bacteria camel dead excite." The examinee is to name the letter trend, which, in this item, of course, is that the initial letters are in alphabetical order from "a" to "e." In the Correlate Completion II test, an illustrative item reads: "am ma not ton tool ———"; what word should come next? Here it is not word meaning that is important but letter sequences. In the Seeing Trends II test, likewise, the word meanings are of no significance. Presumably, an analogies test utilizing letters only would do as well as a measure of this factor.

In the row of Table **1** pertaining to patterns or systems, we have three factors, but they are much more disparate in kind than usual in this table. The clearest example of an eduction-of-patterns factor is in the middle column. The Circle Reasoning test, adapted from Blakey (**9**), is similar to the Marks test of Thurstone and to the Spatial Reasoning test of the AAF (**21**). In a sequence of symbols the examinee must discover the principle by which certain symbols are marked, then he must mark a new set accordingly. In the Letter Triangle test, the letters are arranged in a different alphabetical pattern in each item. The examinee must discover the pattern and show this by filling a blank with a letter.

Under the figural category we find the factor of *spatial orientation*, a well-known space factor. It is best defined as the ability to become aware of the spatial order or arrangement of objects perceived visually.

Until the system of cognition factors was conceived, the writer had thought of *spatial orientation* as a purely perceptual ability rather than intellectual.[5] Its place in the system is regarded as tentative. We may yet find another seeing-patterns factor in which figural properties play a more obvious role than they do in the

[5] A perceptual factor is distinguished from an intellectual factor by the fact that no symbolic activity is clearly involved.

factor of *spatial orientation*. In a real sense, an orientation within a field of perceived objects is a pattern or system, where spatial arrangement, which includes the viewer, is the principle. Shapes and sizes of objects, which play a more obvious role in the case of the other figural factors, are of more indirect significance in *spatial orientation*.

Under the conceptual category we find a factor that has been most difficult to define. The best conception of it is that it represents an ability to define or structure problems. It has been a most consistent component of arithmetic-reasoning tests, but since such tests are psychologically complex, it has been difficult to determine just what aspect of solving problems of this type is the significant feature that requires the ability called *general reasoning*. By elimination of many rival hypotheses, it is now rather clear that the factor pertains to the comprehension of the structure of a problem, at least of the arithmetical variety (**19**). Since such a structure is conceptual, the factor logically belongs in the column where it is placed in Table 1. The Ship Destination test is a special type of arithmetical-reasoning test, which seems to come closer than any other to being a pure measure of the factor.

In the next row of Table 1, for the discovery of problems, there is only one factor—*sensitivity to problems*, which is in the conceptual column. The appearance of this factor parallel to *general reasoning* in the row preceding, emphasizes the well-known observation that it is one thing to be aware that a problem exists and another thing to be aware of the nature of the problem. The titles of the tests are quite descriptive. A sample item from the test Seeing Problems asks the examinee to list as many as five problems in connection with a common object like a candle. The test Seeing Deficiencies presents in each item the general plan for solving a given problem, but the plan raises some new problems. What are those problems?

Whether we shall ever find parallel factors for seeing problems or deficiencies of figural and structural types remains to be seen. Problems of a figural type are faced in aesthetic pursuits such as painting and architecture. Problems of a structural type might be faced in connection with spelling or the development of language. Tests pertaining to the seeing of problems have thus far provided no figural or structural bases for problems. It should be relatively easy to test the hypothesis that such factors exist. If they do exist, their possible implications for everyday performance need further study.

In the investigation of planning abilities (**14**, **15**), two parallel factors were found—*perceptual foresight* and *conceptual foresight* where one was expected. The Competitive Planning test was originally designed by the AAF psychologists as a test of foresight and planning (**21**). It requires the examinee to imagine that he is playing the game of completing squares by drawing lines. He plays for the two opponents and in each item he has to tell the maximum number of squares each opponent can complete under the rules of the game. The Route Planning test, another AAF product, is a type of maze problem. The examinee must say which of alternative points will have to be passed through in going from the starting point to the goal. In both tests, perceived layouts are used.

The test Pertinent Questions presents in each item a need for a decision and the examinee is asked to state what facts he should consider

in reaching a decision. For example, a new graduate is offered positions in two different cities. What should be the deciding considerations? In the Alternate Methods test, a practical problem is given, with available objects that may be used. The examinee is to give several alternative solutions that he considers most adequate.

Porteus has maintained that his series of maze tests measure foresight. He can well claim support from the factor-analysis results just mentioned. The type of foresight measured by maze tests, however, is of a concrete variety. This ability may be important for the architect, the engineer, and the industrial-layout planner. It may not be found related to the abstract type of planning that we find in the political strategist and the policy maker. So far as our results go, the maze test should by no means be offered as a test of general intelligence. This statement might need modification, however, after the maze test is factor analyzed in a population of lower general intellectual level (where general intelligence is defined operationally as an average of all intellectual abilities). In a population of "high-level personnel," we can say that a maze test measures most strongly the factor of *perceptual foresight* and, incidentally, to some degree the factors of *visualization* and *adaptive flexibility* (**15**).

The appearance of a factor called *penetration* in the last column of Table 1, along with *conceptual foresight*, calls for comment. A factor of penetration was hypothesized in the first analysis of creative abilities and was not found (**31**). An unidentified factor found there might well have been *penetration*. A factor has been so identified in a more recent analysis that emphasized creative ability tests (**20**). It is strongly loaded on a test called Social Institutions, which asks what is wrong with well-known institutions such as tipping. It was designed as a test of *sensitivity to problems*, and it has consistently had a loading on that factor. In the first creativity analysis, two scores were based upon this test; one being the total number of low-quality or obvious defects and the other was the total number of high-quality or "penetrating" defects—defects that can be seen only by the far-sighted person. As a matter of fact, the two scores had much to do with effecting a separation of the seeing-problems tests into two groups, one of which might have been identified as the *penetration* factor.

It is quite possible that the factor of *penetration* and the factor of *conceptual foresight* are one and the same. They came out in two different analyses that had no crucial tests in common. It would be a good hypothesis that they are identical and a good prediction would be that if the four tests listed in Table 1 were analyzed in the same battery they would define a single factor, not two.

There is the apparent possibility for the existence of a foresight factor involving structural arrangements, but the scope and usefulness of such a factor would seem to be questionable.

Production Factors—Convergent Thinking

The second large group of thinking factors has to do with the production of some end result. After one has comprehended the situation, or the significant aspects of it at the moment, usually something needs to be done to it or about it. In the analogies test, for example, having seen the relation between the first pair of elements of an item we must then

find a correlate to complete another pair. Having understood a problem, we must take further steps to solve it.

Like the cognition factors, the production factors show some promise of falling under the rubrics of figural, structural, and conceptual, but here the picture is less complete. The kinds of things produced are more numerous than the kinds discovered. There are no identities of things in the two lists, but there are a few parallels or relationships. For example, corresponding to the comprehension of words, there are factors concerned with the production of words; corresponding to the discovery of classes there is the act of naming; corresponding to the discovery of relations there is the production of correlates; and corresponding to the discovery of systems there is the production of order. But with these few instances, the connections and parallels seem to end.

It was announced earlier that the production factors fall into two groups—convergent-thinking factors and divergent-thinking factors. Such a distinction seems not to have been emphasized in prior literature on thinking. In the case of some of the production factors, the distinction is not complete, but in most cases it is striking.

In convergent thinking, there is usually one conclusion or answer that is regarded as unique, and thinking is channeled or controlled in the direction of that answer. In tests of the convergent-thinking factors, there is one keyed answer to each item. Multiple-choice tests are well adapted to the measurement of these abilities. In divergent thinking, on the other hand, there is much searching or going off in various directions. This is most clearly seen when there is no unique conclusion. For the measurement of such abilities, completion tests are almost a necessity. The distinction is not so clear in some problem-solving tests, in which there must be and usually is some divergent thinking or search as well as ultimate convergence toward the solution. But the processes are logically and operationally separable, even in such activities.

In Table 2 we have those production factors identified as dealing with convergent thinking. There are five potential triads of factors, depending upon the kind of result produced—names, correlates, orders, changes, or unique conclusions. In two cases structural-type tests have figured in factors, thus a three-column matrix has been again adopted.

In the first row are factors having to do with the production of names. The two factors there are again contrasted in terms of the concrete-abstract dichotomy. They differ, also, by the fact that the one has to do with the naming of particulars while the other has to do with the naming of classes. French (**8**) lists a factor of naming, which has been called *object naming* here to distinguish it from the factor of *abstraction naming*, which was just recently discovered. The appearance of a test of Color Naming under the rubric of "figural" calls for broadening the conception of this class to recognize color as a figural property. Classes of objects distinguished for their structural properties are evidently not very common. If good examples can be found, we may find a third naming factor. In the name of the factor of *abstraction naming*, the term "abstraction" may prove to be too comprehensive. The two illustrative tests mentioned might suggest that the ability is restricted to the naming of classes. The results show that it is actually broader than that, since

TABLE 2

PRODUCTION FACTORS—CONVERGENT THINKING

Type of result produced	Type of Content		
	Figural	Structural	Conceptual
Names	*Object naming* Form Naming Color Naming		*Abstraction naming* Picture-Group Naming Word-Group Naming
Correlates	*Eduction of correlates* Correlate Completion Figure Analogies Completion		
Orders			*Ordering* Picture Arrangement Sentence Order
Changes	*Visualization* Spatial Visualization Punched Holes		*Redefinition* Gestalt Transformation Object Synthesis
Unique conclusions	*Symbol substitution* Sign Changes Form Reasoning	*Numerical facility* Numerical Operations	*Symbol manipulation* Symbol Manipulation Sign Changes II

it pertains to the naming of relations also, in other tests.

With three factors having to do with the seeing of relationships, we might well expect three corresponding factors concerned with the eduction of correlates. As a matter of fact, the project has for some time anticipated at least two such factors, perceptual and conceptual, and has designed tests that were expected to effect the expected separation. To this date, only one eduction-of-correlates factor has been clearly indicated, and both figural and structural tests have loadings on it. The Verbal Analogies Completion test, which we hoped would help to distinguish a conceptual-correlates factor, turned out to be a test of *expressional fluency*. Evidently the eduction-of-correlates aspect of the test was made so easy that little variance in this ability, if it is separate, was manifested. On the other hand, having educed the correlate, thinking of the needed word provided the chief basis for individual differences in scores, and hence the loading on *expressional fluency*. It can be predicted that with the appropriate tests, three eduction-of-correlates factors will become evident. Because of the difficulty of separating them, it can be predicted that the intercorrelations of these three factors will be found to be substantial.

In the investigation of planning abilities it was hypothesized that there would be an ability to see or to appreciate order or the lack of it, as a feature of preparation for planning. It was also hypothesized that there would be an ability to produce order among objects, ideas, or events, in the production of a plan. A single ordering factor was found. Since the three tests designed to measure sensi-

tivity to order had low and insignificant loadings on the factor, while the three designed to measure the production of order had significant and even substantial loadings, the factor seems to belong among the production factors. The Picture Arrangement test presents a four-part cartoon strip in which the parts are out of correct temporal order. The examinee has to state the best order. The Sentence Order test presents in each item three sentences, each stating an event, the examinee being told to rearrange them.

It remains to be seen whether ordering in terms of figural and structural properties will call for additional ordering factors to help complete the matrix of Table 2. Figural ordering may be a significant aspect of pictorial art. It is not so easy to see where a structural ordering would be of consequence.

In the next row of Table 2 we find the factor of *visualization*, which has been known for some time, and the factor of *redefinition*, which was found originally in the first creativity analysis (**31**). The thing produced in both instances is some kind of change or rearrangement or shift. The Spatial Visualization test is Part VI of the *Guilford-Zimmerman Aptitude Survey*. In each item certain movements of a pictured alarm clock are indicated and the examinee is to select the view that would be seen after the movements. The Thurstone Punched Holes test shows a paper being folded and a hole or holes then cut out. The examinee is to tell how the paper would look after unfolding.

The *redefinition* factor involves shifts of meaning or use of objects or parts of objects. The test Gestalt Transformation asks such questions as: With which of the following objects could one best start a fire: A. fountain pen, B. onion, C. pocket watch, D. light bulb, E. bowling ball? The keyed answer is C, since the crystal can be transformed from a face cover to a condensing lens. The Object Synthesis test asks such questions as: Given pliers and a shoestring, what could you make? A good answer would be "pendulum" or "plumb bob." In either case the objects play new roles in the combination.

The last row of factors in Table 2 presents an interesting triad. Although there are one or two questions that can be raised about their placement, to be mentioned later, it is quite clear that they all involve rigorous operations with symbols leading to unique conclusions. The factor of *numerical facility* is the very well-known ability to operate with numbers, where both speed and accuracy are significant. The two new factors, *symbol substitution* and *symbol manipulation*, were regarded as one factor until recently. In one analysis the factor looked like a substitution ability and in another analysis it looked like a manipulation ability. In a recent analysis (**20**) the two were found to be separate.

To distinguish these factors, we must consider the different kinds of tests that represent the two. In Sign Changes, the examinee is told before each block of items what interchanges to make in algebraic signs, e.g., "replace − with ×" and "replace + with −." He applies the new rules to several simple equations such as "3−6=?" and "6+2=?." In the Form Reasoning test, equations are stated in the form of combinations of simple geometric forms. Some definitions are first given, stating that a combination of two forms, such as a star and a circle, can be replaced by another single form, a square. With these substitutions of single forms for pairs, combinations

greater than pairs must be reduced to single symbols, taking each pair in turn.

It is difficult to accept fully the placement of *symbol substitution* in the figural column. If all tests loaded on it were like Form Reasoning, where the rigorous definitions and operations are all in terms of figures, the placement would be quite reasonable. But certain features of the Sign Changes test suggest that it is not figural properties, as such, that are important. They may serve merely to identify the symbols. In the Sign Changes test it is the operation that the symbol stands for that is important.

The Sign Changes test was originally designed as a flexibility test; the Form Reasoning test was not. In both, the readiness to switch the meaning or significance of symbols is the obvious peculiar feature. Perhaps the emphasis should be placed on the word "switch." It may be that this factor will eventually be placed in the family of flexibility factors, which appears in Table 3. There is no evidence against the hypothesis that *symbol substitution* is the same as the present factor of *adaptive flexibility*, represented particularly by the Match Problems test. As a matter of fact, Sign Changes had a significant loading on *adaptive flexibility* in the creativity analysis (**31**). Form Reasoning has

TABLE 3

PRODUCTION FACTORS—DIVERGENT THINKING

Type of result produced	Type of Content		
	Figural	Structural	Conceptual
Words		*Word fluency* Prefixes Anagrams	*Associational fluency* Controlled Associations II Associations III
Ideas			*Ideational fluency* Plot Titles Consequences
Expressions			*Expressional fluency* Vocabulary Completion Similes
Shifts	*Flexibility of closure* Hidden Pictures Gottschaldt A	*Adaptive flexibility* Match Problems Planning Air Maneuvers	*Spontaneous flexibility* Brick Uses Unusual Uses
Novel responses			*Originality* Plot Titles (cleverness) Symbol Production
Details	*Elaboration** Planning Elaboration Figure Production		*Elaboration** Planning Elaboration Figure Production

* At present regarded as the same factor, but future results may indicate two separate factors.

never had an opportunity to show such a loading.

Defining the factor of *symbol manipulation* are the two tests Symbol Manipulation and Sign Changes II. Symbol Manipulation provides some simply defined symbols, such as: E means equal to; NG means not greater than. Each item then provides a statement such as: xEy and yNGz; which of the following statements can logically be made: xSz, xNGz, etc. This test was designed originally for the factor of *logical evaluation* (see Table 4), and has usually shown some relationship to that factor, but it also helps to define the factor of *symbol manipulation*.

The test Sign Changes II presents simple "equations" such as $1+2=4\times1$, the two sides of which are not actually equal as the statement stands. The examinee is to say what interchange of algebraic signs will make the equation correct. In the illustration just given, if $\times$ and $-$ are interchanged the equation will balance.

From these two tests alone, it is not easy to see exactly what kind of ability is involved in common. One clue may be that both tests involve equations. A third test with a significant loading in one analysis is a number-series test. This test does not involve equations. In one analysis the *numerical-facility* factor was distinct from *symbol manipulation*, consequently we cannot identify the latter with the former. Further intensive work is obviously needed in the area of these factors. Abilities that may be of some significance for success in mathematics may be found here.

Production Factors—Divergent Thinking

The divergent-thinking factors are arranged in a matrix in Table 3, with the three column categories that have now become familiar. Here there are more vacancies to be filled, if the system is indeed as applicable as it promises to be.

In the first three rows of the table we find the four well-established fluency factors. In the first row are the

TABLE 4

EVALUATION FACTORS

Type of Content		
Figural	Structural	Conceptual
(*Perceptual evaluation*)* Ratio Estimation Figure Estimation *Length estimation* Pattern Assembly Shorter Path		*Logical evaluation* Logical Reasoning Inferences *Experiential evaluation* Unusual Details *Judgment* Practical Judgment Practical Estimation
	Speed of judgment Color-Form Sort Time Social Judgments Time	

* Probably a composite of factors, including *length estimation*.

two fluency factors having to do with the production of single words. In the case of the factor of *word fluency*, meaning is of no importance. The usual tests of this factor merely specify that the words shall begin or end with a specified letter, prefix, or suffix. Only such structural requirements are to be met. The examinee need not even know the meanings of the words he gives. In the case of *associational fluency*, however, meaning is an essential requirement. The words given must be synonyms, as in Controlled Associations II, or must be related in some meaningful way to stimulus words or ideas. In Controlled Associations II, the examinee gives as many as three synonyms to each stimulus word. In Associations III, two words are given, differing in meaning, and the examinee must give one word that is a synonym to both. For example, the word "lie" would be given as a synonym to both "recline" and "deceive."

It does not seem very likely that an ability will be found for the first cell in Row 1 of the table. This would call for the production of words satisfying specified figural requirements. Yet, tasks can be thought of to meet this case, for example, the writing of headlines, the production of esthetic effects with words, and so on. It does not seem likely, however, that there should have developed in human makeup a unitary ability of this kind.

The second row of the table offers some interesting possibilities. The speed of calling up ideas expressible in verbal form can be tested by different kinds of tasks. The two examples of tests given were designed for the study of creativity. The Plot Titles test of fluency is scored by the total number of low-quality titles that can be suggested for a short story plot in a given time. The Consequences test is scored similarly, but the responses are consequences foreseen as a result of some drastic change, such as everyone going blind.

It can well be questioned whether fluency of verbal responses of such kinds is strongly related to fluency of ideas of a mechanical, or musical, or pictorial kind. Fluency tests have been commonly cast in verbal form. Fluency in the production of figures and fluency in the production of things distinguished by their structural properties may well be separate factors, both distinct from the *ideational-fluency* factor now known. The exploration of such possibilities would seem to be a fruitful route to take in the investigation of creativity.

The separateness of the factor *expressional fluency* from *ideational fluency* indicates that the ability to have ideas and the ability to put them into words are different things. Since the examinee must state verbally his ideas in tests of *ideational fluency*, it might be supposed that his ability to express himself is included or is also being tested. But apparently in such a test the expressional problem is not a serious one. We present other tests in which the idea is given and the examinee must put it into words, usually in more than one way. The expressional problem is then more difficult, the test giving us variance in the expressional factor. In the Vocabulary Completion test, a stimulus word is used in a brief context, enough to indicate its meaning, and the examinee has to give the word. In the Similes test, the examinee must give more than one completion to a simile. In a Verbal Analogies Completion test, which was designed to measure another factor, we found that the leading variance is in the *expressional-fluency* factor.

The only complete triad in Table 3 is a set of flexibility factors, the

best-known of which is *adaptive flexibility.* The three factors involved are not clearly parallel in all respects. They have in common the feature that sudden shifts of activity occur—shift of organization of a figure, shift of set or approach to a problem, or shift of category of responses, respectively. Thurstone discovered the *flexibility-of-closure* factor in his analysis of perception (**28**) and found that the factor had relations indicating its intellectual importance.

The most consistently representative test of the factor of *adaptive flexibility* is the Match Problems test. This test is based upon the old, familiar puzzle or game of removing a specified number of match sticks in order to leave a specified number of squares. In order to measure flexibility, the problem changes drastically from one item to the next, requiring very unusual solutions; solutions such as the average person would not expect. For example, at first the examinee is led to expect that the remaining squares will be of the same size, but there comes an item in which they must be of unequal size. Another item requires that a smaller square be left within a larger one, and so on.

In an unpublished study, a test involving Gottschaldt figures came out as strongly loaded on *adaptive flexibility* as did Match Problems. In the same analysis, a test of Insight Puzzles also had a similar loading. Thus, in this case, a perceptual, a structural, and a conceptual test had strong loadings on the same factor. There is therefore the possibility that *flexibility of closure* and *adaptive flexibility* are one and the same factor and that this factor cuts across all three columns of the matrix. In an analysis where perceptual, structural, and conceptual flexibility tests are all liberally represented, however, it can be predicted that three factors will be found. If so, they are probably substantially intercorrelated. If there are three such factors, the factor of *spontaneous flexibility* would have to be moved to another row of the matrix to be replaced by a conceptual-adaptive-flexibility factor.

The factor of *spontaneous flexibility* has appeared persistently but never with great strength or stability. The Brick Uses test, flexibility score, is the best clue to its nature. This score is the number of *runs* of responses. The examinee is told to name all the uses he can think of for a common brick, in eight minutes. A "run" of responses is a sequence of uses all of the same class, such as the use of bricks as building material or as missiles, and so on. The test Unusual Uses calls for listing several unconventional uses for each of a number of objects, the number given being the score. Since only verbal tests of this factor have been analyzed, nothing can be said regarding the possibility that there are parallel factors involving figural and structural contents.

It is of some interest to attempt to relate *spontaneous flexibility* to other concepts in psychology. Essentially, it appears to be a disposition to avoid repeating one's self. This suggests a relation to Thorndike's concept of refractory phase or to Hull's concept of reaction inhibition. A hypothesis to be tested would be that tests designed to measure individual differences in tendency to show refractory phase of the Thorndikian type and tests to show degree of tendency to reactive inhibition indicate the same factor as do tests of *spontaneous flexibility.*

The results continue to show that *originality* is operationally definable as the likelihood of giving unconventional, clever, or remotely asso-

ciated responses to test items (**30**). It is measurable in terms of number of clever titles given to story plots, clever "punch lines" for cartoons, remote consequences to events, and idiosyncratic word associations. In two analyses there has been opportunity for a cleverness factor to separate off from the rest, but this did not occur. While the factor thus seems to be a rather broad one, it may well be asked whether such a factor, measured only by means of verbal tests, is significantly related to original production in nonverbal activities such as graphic arts, music, or inventive engineering.

We have had only one *originality* test that is at least partly nonverbal—the Symbol Production test. This test was designed for another purpose, namely to test the hypothesis that there is a separate ability to symbolize ideas in terms of simple line drawings. Each item presents a statement, such as "*ring* the *bell*," of which the two italicized words are to be represented by two symbols. The score is the number of nouns and verbs symbolized in the testing time. The test is not entirely nonverbal, of course, although the thing produced is figural. There was a second test (Line Drawing) requiring the production of line symbols for given adjectives in the same battery with the Symbol Production test. These two tests might have given rise to a separate factor, but they did not. Nevertheless, the writer is of the opinion that the problem of whether there are originality factors peculiar to nonverbal areas is still an open one.

The *elaboration* factor is an ability to provide details working toward completion, when a part or an outline is given. The test Planning Elaboration presents the bare outline of a plan to which details must be added to make it effective. In the Figure Production test, a simple line is given, to which the examinee is asked to add lines to complete an object. The score depends upon the amount of detail added.

Here we have a clearly verbal test and a clearly figural test (although a meaningful object is usually produced) both with relation to the same factor. There is still the possibility that there are two (or three) elaboration factors, distinguished in terms of content, with enough relationship between them to cause the factors to appear to be one. It will take a new analysis in which at least three good figure-elaboration tests and three good verbal-elaboration tests (not to forget a triad of structural-elaboration tests, also) should be included to determine how many elaboration factors there are.

Considering the factors in the divergent-thinking category together, it is obvious that the freedom to change direction of thinking varies considerably from one instance to another. Different degrees of situation-imposed restriction are involved. But generally, within whatever limits that are imposed by external restrictions, the need for rejecting or superseding a response and for trying or producing a new one is the common element in this group of factors. There is also a difference in the amount of self-imposed restriction or freedom. This depends upon the individual rather than upon the situation. It is largely in this source of variation that we find the divergent-thinking factors.

Evaluation Factors

Evaluation factors have to do with decisions concerning the goodness, suitability, or effectiveness of the results of thinking. After a discovery is made, after a product is achieved,

is it correct, is it the best that we can do, will it work? This calls for a judgmental step of some kind. It was our hypothesis in the project that the ability to make such decisions will depend upon the area within which the thinking takes place and the criteria on which the decision is based. The results indicate several evaluation factors. They have been placed in the customary three-column matrix in Table **4**, in spite of the fact that none have been found to fit the structural column. In this group of factors there is no good way of distinguishing rows. The domain of evaluation factors has been less well explored than the other intellectual domains.

The least that can be said is that the perceptual-conceptual dichotomy applies in this area of abilities. Although our analysis showed only one factor applying to judgments of figural material, it is likely that in this subarea of evaluation alone there are a number of judgment factors. For this reason the factor of *perceptual evaluation* has been placed in parentheses in Table **4**. For example, a more restricted factor of *length estimation* has been found (**21**). The search for such factors carries us over into the whole realm of psychophysical judgment. It would be difficult to say whether factors of this kind belong under the general heading of thinking or under the heading of perception. In view of the known complexity of psychophysical judgments in general, their place in the intellectual group can be defended.

The best established evaluation factor is that of *logical evaluation*. This is defined as the ability to judge the soundness of conclusions where logical consistency is the criterion. The factor has sometimes been called "deduction," with the belief that it is the ability to draw conclusions logically consistent with premises. If this were the case, the factor would belong with the production-factors group. Most tests in which the factor has been found to be a component are of the true-false or multiple-choice form, in which the examinee is given conclusions; he need not produce them. It is difficult to say whether he actually does produce them for himself first then find them among the answers provided. But whether he does this or not, he must necessarily make a judgment as to the correctness of the answer—his own answer or the ones given him. Even in a completion test, this step would be necessary. It seems preferable, therefore, to call the factor *logical evaluation* and to list it among the evaluation factors.

It was hypothesized that there would be a factor in which evaluation is made on the basis of past experience. Such a factor was found, and it is represented best by the test of Unusual Details. In this test the examinee is asked essentially "What is wrong with this picture," in which there are two features that are incongruous or inconsistent with common experience. In defining this factor, whether the emphasis should be placed upon the supply of past experience or upon an ability to utilize that experience is not known.

The factor called *judgment* is listed with some hesitation. It was found repeatedly, but rather weakly, in AAF research (**21**). It is best represented by a test in which a practical difficulty was described and several alternative solutions are offered. Which one is best, everything considered? In common terminology, the ability might be recognized as wisdom or common sense. In the aptitudes-project research, there is evidence that this AAF *judgment* factor may be the same as the one called

redefinition. If this is the case, it is not easy to say where to place the emphasis in defining the factor.

The factor ***speed of judgment*** was found by Thurstone in his analysis of perceptual abilities (**28**). The speed with which the examinee completes the sorting of objects according to color or form and the speed with which he checks traits that apply to himself are both measures of the factor. It is thus shown as cutting across the three content categories. It might well be classed as a temperament trait rather than an ability.

Memory Factors

There is little doubt about the grouping of the remaining factors under the heading of memory factors. Collecting all such factors from various sources, we find that seven qualify for this category. A recent analysis by Kelley (**27**) has done much to verify and complete the picture for this group. It is possible to organize these factors in the three columns of the now familiar categories as to content, and in three rows as to the kind of thing or aspect involved (see Table 5). The titles of the tests representing each factor are usually quite descriptive.

The best-known of the memory factors is ***rote memory***; the ability to learn and to remember things associated, where meaning is of little or no importance. In the AAF research this factor was called "associative memory" for the reason that paired-associate learning was typical of the tests of it. There was a need, also, of distinguishing it from the factor of ***visual memory***, where sheer content is important rather than associative connections between contents. Since Kelley (**27**) has demonstrated another associative-memory factor in the form of ***meaningful memory***, however, it seems best to return to the name of ***rote memory***. The placement of both in an associative row of the matrix indicates their common associative property. The vacancy under the figural heading in this row calls for the hypothesis that there is an undiscovered factor pertaining to the learning of associative connections between figural contents.

TABLE 5
A Matrix of Memory Factors

Thing or aspect remembered	Type of Content		
	Figural	Structural	Conceptual
Associative connections		*Rote memory* Word-Number Color-Word	*Meaningful memory* Sentence Completion Related Words
Content	*Visual memory* Reproduction of Designs Map Memory *Auditory memory* Musical memory Rhythm		*Memory for ideas* Memory for Ideas Limericks
Span		*Memory span* Letter Span Digit Span	*Integration I* Signal Interpretation Combat Planes

The factor of *visual memory* has been known for some time (**21**). The factor may be regarded as a rather photographic-memory ability. Some individuals are recognized as standing out in this respect, for example certain police officers who remember faces and motor-vehicle license numbers remarkably well. In tests, the evidence of remembering of this type may be in the form of reproductions (Reproduction of Designs test), or recognition (an AAF Map Memory test), or verbal descriptions (another AAF Map Memory test).

The listing of a factor with the name of *auditory memory* represents in part the writer's somewhat risky hypothesis. It is based upon a factor found by Karlin (**26**) in tests of musical memory (for melody and rhythm). French (**4**) called it "musical memory," which is the cautious thing to do. The name "auditory memory" used here implies some confidence in the prediction that when nonmusical auditory-memory tests are included with musical-memory tests in the same analysis, the same factor will apply to both.

AAF research results hinted at the existence of a content-memory or substance-memory factor but did not demonstrate it. Kelley's results give evidence for such a factor. It is the memory for ideas, which are probably not expressed verbatim in recall tests. Further support for this factor is desirable. The hypothesis that there is a "content" factor in the structural column is still to be investigated. It is not easy to say what this would be like. The memory for a route might qualify.

Memory-span tests, composed of digits and letters have in common a *memory-span* factor. This factor belongs in the structural column. Incidentally, it is interesting that memory-span tests have been rather popular components of general-intelligence scales. It turns out that they measure primarily a rather special kind of memory ability whose social importance cannot be very great. Telephone operators come to mind first in this connection. A general remark may be made, prompted by the emphasis upon memory-span tests as measures of intelligence, that although many tests correlate highly with chronological age, this does not ensure that they measure any very significant aspect of intelligence.

In the conceptual column, *Integration I*, a factor found in AAF research, is proposed as a memory-span factor. The tests Signal Interpretation and Combat Planes require the examinee to keep in mind a relatively large number of detailed rules for success in them. Kelley (**27**) had one span test in which the content was in the nature of lists of tasks to be done, the length varying as in digit and letter-span tests. It came out with those other span tests on his *memory-span* factor. It can be predicted that if there were other idea-span tests, and perhaps some *Integration-I* tests in the battery, two span factors would be found.[6] The span factors are probably significantly correlated. The vacant cell in Row 3 of Table 4 suggests that the way is open for someone to see whether a third memory-span factor will be found where the contents are figural.

To digress somewhat from an account of the factors, it may be pointed out that the fact that there are several distinct memory abilities may explain some of the phenomena observed in memory experiments, particularly where results are discordant. Results from memory ex-

[6] Another hypothesis is tenable with regard to *Integration I*, however. It might be identical with the factor *memory for ideas*.

periments may differ markedly, sometimes, depending upon the kind of material and the thing or aspect emphasized. For example, the relative strength of backward vs. forward associations differs when the material is composed of visual forms or is composed of syllables. In transfer experiments, in view of the different abilities involved, it should not be surprising that transfers of gains in memorizing skills should be so limited. It would be interesting to test the hypothesis that transfer will be relatively greater between tasks that depend upon the same memory factor or upon the more strongly correlated factors. The same hypothesis could be stated with respect to thinking factors and other ability factors generally.

Discussion

The account of the known intellectual factors and the system into which they seem to fall calls for the discussion of some general questions. There are implications for factor theory and for its application to psychological research in general. There are implications for general psychological theory and for the practices of intelligence testing.

Implications for Factor Theory and Factor Analysis

A theory or a method should be judged by its fruits. If the results that have been reported here contribute to psychological understanding and, through that, to useful psychological practice, factor analysis has passed this kind of test. The mathematical model that has been applied, which conceives of individual differences in intellectual performances as being represented by a coordinate system of *n* dimensions, has served certain purposes. While it may be shown at some future time that the model is not the best that could be applied, its power to generate new psychological ideas and to extend considerably the conception of the realm of intellect has been demonstrated.

The average reader will no doubt be surprised by the large number of dimensions that seem to be required to encompass the range of intellectual aspects of human nature. Some 40 factors are reported as being known and a great many additional unknown factors are forecast. This would seem to go against the scientific urge for parsimony.

The principle of parsimony has led us in the past to the extreme of one intellectual dimension, which everyone should now regard as going too far in that direction. There is actually no fixed criterion for the satisfaction of the principle of parsimony. In science we can satisfy the principle to some degree whenever the number of concepts is smaller than the number of phenomena observed. Forty, sixty, or even a hundred factors would certainly be a smaller number of concepts than the number of possible tests or the number of observable types of activities of an intellectual character. In this sense the principle of parsimony has been satisfied.

The number of the factors is less unattractive when we find that they can be subsumed within a system that is describable by a smaller number of categories or principles, as we have seen in the matrices of Tables 1–5. Some readers will ask whether, since there are many probable intercorrelations among the factors, a small set of second-order factors will not suffice. Granting that we can make sufficiently accurate estimates of the intercorrelations among the factors, which the writer doubts that we can do at present, to use only sec-

ond-order-factor concepts would lose information. This follows from the fact that where *n* linearly independent dimensions are necessary to describe a domain geometrically, no one dimension can be entirely accounted for by combinations of the others.

It may be asked whether some of the factors listed are not really specific factors rather than common factors. This is a legitimate question. It is not uncommon experience in factor analysis to find what was formerly regarded as a single common factor appears later to split up into two or more factors. The "splitting up" description is not completely accurate. It applies best to the fact that a group of tests having a "factor" in common later divide into two or more groups each defining its own common factor. In clear thinking about this phenomenon, we must keep in mind the distinction between "factor" as a mathematical concept and "factor" as a psychological concept. The immediate results of a factor analysis are in terms of mathematical factors. Whether each mathematical factor represents a single psychological factor or a combination of psychological factors has to be determined by interpretation and by further experimental work applied to the designing of new factor analyses. Eventually we reach the stage where further efforts to "split" a factor fail. Whether this has brought us to a specific factor in any particular case can be decided on the basis of a single criterion. Are the tests defining this factor essentially just different forms of the same test? This cannot always be decided with certainty, but there is usually little difficulty in doing so. If we suspect that any factor is a specific, a new analysis that includes more obviously different tests, but tests that should measure the same *common* factor, should be done.

Skepticism was expressed above concerning the operation of estimating factor intercorrelations. This is a somewhat complicated problem for which there is as yet no good solution. The common procedure in vogue at the present time for estimating factor intercorrelations is to do an oblique rotation of axes, locate the primary axes and determine the cosines of their angles of separation. The writer has preferred orthogonal rotations for several reasons. Briefly, any particular oblique solution to a factor problem is a function of several nonpsychological circumstances. For one thing, it depends upon the kind of population tested. This is not so serious, but we should probably have a different set of factor intercorrelations for each age group, educational level, cultural milieu, etc., and for combinations of these. This lack of invariance precludes making any very general statements regarding the psychological interdependencies of factors.

A more serious matter is that oblique solutions depend upon the population of tests that we factor analyze. This is not merely a sampling problem, for the collection of tests in a battery is never a randomly selected one, and should certainly not be. Much of this difficulty hinges on inadequacies of test construction and test administration. Rarely do we succeed well enough, either by test construction or by test administration, in exerting the experimental controls it would take to come out with a score that is a pure measure of a factor. If two factors happen to be commonly loaded in the tests that define both of them, it would give the appearance of a factor intercorrelation whether there was genuine correlation or not. This

kind of result is not uncommon. Until we succeed in exerting better experimental controls in testing, we shall not have a very good basis for estimating factor intercorrelations, even for a specified population of examinees.

The question always comes up regarding the origins of factors; are they inherited or are they acquired, to use the common, loose expression of this question. The reply is that factor analysis alone cannot answer this question. So far as factor analysis is concerned, the factors could all be hereditary in origin, or all environmental, all some weighted combination of both heredity and environment, or some due to the one and some to the other source. It will take experimental work of the usual types to answer this question. But one thing is clear. The question "Is intelligence inherited or is it acquired" makes less sense than it ever did. Such a question must be asked regarding each and every factor. Ferguson (**4**) has expressed the interesting hypothesis that factors are a consequence of the principles of transfer of learning. Many of them may be, to a large extent. The Ferguson hypothesis is akin to a similar one expressed earlier in this paper.

In connection with origins of factors, there is also the question of when in child development the factors make their appearances. To the extent that factors are developed by experience, they would appear at such ages as the effects of experience have sufficiently crystallized. To the extent that heredity is chiefly responsible for the differentiation of factors, their appearances should be detectable when maturation effects their differentiation. In either case, the answer is to be determined by experimental testing and factor analysis at all age levels at which suitable tests can be administered. Such analyses should be done in populations very homogeneous with respect to age and other features. It can be predicted that the structure of the intellectual factors for children will be found simpler than that for adults. It can also be hypothesized that the structure for generally superior adults will be found more complex than for generally inferior adults.

Implications for Psychological Theory

It was suggested earlier that although psychological factors are variables among individual differences they also indicate psychological functions within individuals. It is therefore in order to take the factors seriously as starting points for psychological theory.

There has never been developed a comprehensive theory of thinking. We have been short of the essential concepts needed in the construction of such a theory. In view of the great variety of thinking abilities (and functions) revealed by factor analysis, the time-honored concepts of reasoning, induction, deduction, and the like appear even more inadequate than before. It seems to be of little value to attempt to relate the factors to those categories. The factors, instead, have generated their own categories, which have been already presented. They are essentially operational concepts, since, like factors, they refer back to the kinds of tests from which factor definitions were inferred.

Although the general picture of the thinking factors is not yet sufficiently complete or certain to suggest an obvious, general theory of thinking, the kind of theory that they will eventually generate can be seen.

It is fairly well agreed that thinking is symbolic behavior. It is not surprising, then, that certain factors

have to do with symbols, as such, and with their utilization and manipulation. Of all the kinds of symbols available to humans in almost any culture, words and numbers are among those of greatest importance. The factors reflect these facts.

In the operations of thinking, of realistic thinking, in particular, the factors indicate the important steps or processes of discovery, production, and evaluation, often occurring roughly in that temporal order. Divergent thinking may come into the picture along with these other phases, and auxiliary to them, particularly when they proceed with some difficulty. Some divergent-thinking processes are also likely to occur in nonrealistic thinking, when one is simply free to do so and finds it rewarding. Since realistic thinking is usually convergent, particularly when there is one right answer, at times there may be conflicting divergent-convergent tendencies, a phenomenon that has not been reported, to the knowledge of the writer.

Quite generally, it seems, the thinking processes of a person may proceed more or less ably depending upon the kind of content with which he is involved—perceived figures, recognized structures, or conceived meanings. The distinction that has sometimes been made between concrete thinking and abstract thinking has foreshadowed the major distinction here; the distinction between figural factors and conceptual factors. The appearance of the third category—structural—came as a surprise. If it turns out to be important, we have several interesting implications.

One practical implication of the structural category is that tests based upon letter material and the like may be of limited significance, if in reality we are interested in predicting behavior that depends upon factors in the figural or conceptual columns. A more important implication has to do with the fact that there is a shortage of known factors in the structural column. A rather direct reason for this may be that there has been a bias toward figural and verbal test material, with an unfortunate slighting of structural material. This would not be so unfortunate if it turns out that in our civilization not many such factors exist, or if they do exist they are of relatively little social importance. It may be that there is actually more structural-type thinking going on than we realize and that both psychologists and educators have failed properly to recognize it. In a highly technical age, such thinking would seem to be important. We might well ask ourselves whether we have overlooked something of importance in this general area.

The headings of rows in Tables 1–3 present an unusual list of concepts, which appear to be more epistemological than psychological. Is this possibly the kind of concepts that we have needed? It may be possible to give some of them more psychological terminology later, but at present they refer to the kinds of things that we can know and can produce. If such terminology describes behavior in a significant and useful manner, it should be welcomed and its worth should be recognized. One implication is that the lists seem to be open to new additions. Consideration of what categories might be added to the lists might turn up some new fruitful hypotheses regarding unknown factors and functions.

The subject of problem solving has come into considerable prominence in recent years. The picture of the thinking factors has important

implications for problem solving. We find that there is no one factor that can be called problem solving. This is significant. Problem solving is usually a complicated process. It is clearly indicated that we should stop looking for any one function or process that is the *sine qua non* of all problem solving. As the writer has pointed out elsewhere, many factors, including perceptual factors as well as thinking factors, may be called into play, depending upon the nature of the problem (**12**).

In the list of thinking factors we find one factor having to do with the ability to recognize that a problem exists and another factor that pertains to the diagnosis of the problem. The degree of generality of either factor is still to be determined. So far as we know now, either may be restricted to a relatively narrow category of problems. The next steps in the attack on problem solving should be to make a survey of the variety of problems that are common and to attempt to write specifications regarding the factorial abilities that play significant roles in the solution of each type of problem. We should then test these hypotheses by experimental and factor-analytic procedures.

At the beginning of the aptitudes-project investigation of creativity it was hypothesized that certain special, creative factors would be found, a few of them being then already known, some not. The results have supported most of the hypothesized factors but not all (**20, 31**). Because these factors were investigated within the arbitrarily designated domain of creativity, there has been a tendency to think of them as being the exclusive creative factors. This conception is not fully correct. Creative thinking, like problem solving (they may actually overlap in many cases), depends upon different combinations of factors, and the combination of factors significant to the task will vary from time to time. The problem confronting us here, as with problem solving, is to recognize the main categories of creative production and to seek the significant combinations of factors involved in them. Although certain factors such as *ideational fluency* and *originality* will carry relatively more weight, other factors not obviously creative may often be significant, as when an invention depends upon thinking by analogy or upon visualization.

Thinking has many connections with learning, and hence the thinking factors are of some importance in learning investigations and learning theory. Thinking is sometimes regarded as a form of learning, for while we think we usually learn. Another view of the connection is that thinking contributes to learning. The latter view is more productive of approaches to investigation of the role of factors in learning. It is not enough to conclude that thinking contributes to learning or even to state and to test this as a general hypothesis. The questions raised here should be "Where and how does factor X contribute to learning?" just as it was asked in the preceding paragraphs where and how each factor contributes to problem solving and creative activity. Since problem solving and creative activity are properly regarded as instances of learning, we need only generalize the question to make it apply to all learning. Fleishman and Hempel (**5, 6, 7**) have already provided some excellent demonstrations of the roles of factors at different stages in the learning process for certain psychomotor tasks. This type of investigation should be applied more generally. Certainly we should have

outgrown the glib definition that "intelligence is learning ability."

The distinction between associative and content-memory factors reminds us that not enough attention is generally paid to the same distinction in studies of learning and memory. Learning theory has restricted itself almost entirely to the formation and retention of associative connections, leaving out of account the learning of substance.

Speaking of learning suggests the practical operation of education. At some future time factors should have much effect upon educational practices, in addition to those effects having to do with assessment. If training and experience have much to do with the development of the factors, it is important to know the factors and to determine the procedures whereby their development can be promoted by education.

There are many possible relationships of the intellectual factors to pathology. Defects of memory and thinking are common occurrences in connection with intellectual losses that are associated with organic and functional pathologies. If we find by observation and by experimental study that defects tend to be along the lines of the intellectual factors, we have another source of evidence for the validity of the factors as functional unities. In practice, the use of measures of the factors may be helpful in providing more accurate and meaningful assessment of intellectual losses. Losses described in terms of the factor concepts may help in understanding the types of pathology, and in providing better definitions and diagnostic criteria.

Intelligence and Intelligence Tests

A treatment of the factors of intellect would be incomplete without considering their implications for the concept of intelligence and for the present and future of intelligence testing. Is the concept of intelligence still useful? What is the nature of current intelligence tests in terms of factors? What should the future trends in intelligence testing be?

As to general terminology, the term "intellect" can be meaningfully defined as the system of thinking and memory factors, functions, or processes. The term "intelligence" has never been uniquely or satisfactorily defined. Factor analysis has fairly well demonstrated that it is not a unique, unitary phenomenon. A "general factor," found by whatever method, is not invariant from one analysis to another and hence fails to qualify as a unity, independent of research circumstances, as Vernon has well stated (**29**). The methods of multiple-factor analysis, which have been chiefly responsible for discovering the factors listed above, do not find a general psychological factor at the first-order level and they find no second-order factor that can properly lay claim to the title of "intelligence."

The term "intelligence" is useful, none the less. But it should be used in a semipopular, technological sense. It is convenient to have such a term, even though it is one of the many rather shifty concepts we have in applied psychology. It would be very desirable, for purposes of communication and understanding, to specify a number of intelligences—intelligence A, intelligence B, and so on. This could be done in terms of the combinations of certain intellectual factors and their weightings in the combinations.

We have such combinations now in connection with the intelligence tests and scales in common use. Let us consider what kind of combinations we have in two of the most used in-

telligence scales. A really good factor analysis of the Stanford Revision of the Binet scale would be rather difficult, and cannot be done satisfactorily without adding to the analyzed battery a liberal number of reference tests. This has never been done. The best analyses that we have were done by Jones (**24**, **25**), who found ten factors among 30 selected items. His resulting picture is not clear because among the 30 items were essentially alternate forms of tests (at different age levels) and no outside reference tests were used. A fully satisfactory analysis of the Stanford-Binet items would undoubtedly reveal more than ten factors present.

It should be noted that when so many factors are present, a composite score based upon all the items can measure each component only to a small degree, if they are nearly equally weighted in the composite. It can also be predicted that the factorial composition of the Binet IQ will be found to vary somewhat from one age level to another. This feature may contribute to a small extent to obtained changes in IQ where substantial age differences are involved.

As it actually happens, a Stanford-Binet IQ, or any IQ from a test whose components are predominantly verbal, is a total score heavily dominated by the *verbal-comprehension* factor. This leaves the other factors with little or no effective voice in the composite, even though they are represented in the scale. In nonverbal intelligence tests, there is likely to be less domination by any one factor, but the nature of the composite varies considerably from battery to battery.

Analyses of the components of the Wechsler-Bellevue scale have also been generally inadequate. The most adequate analysis has been done by Davis (**3**), who utilized a number of reference tests from outside the Wechsler battery. He found nine common factors, six of which are probably to be identified with factors in the intellectual list. Where standard tests of intelligence are widely used, it becomes increasingly important to attempt to write the specifications for their total scores as well as their part scores, so that obtained scores of individuals may be most meaningfully interpreted.

Intelligence tests will probably continue to be used for some time to come much as they are. In order to use them most wisely and to extract the greatest amount of information from their scores, the specification of such scores in terms of known factors is one important improvement that could be made. The other great step toward improvement in intelligence testing would be to emphasize more than at present some of the socially important factors that have to do with productive thinking. The knowledge of the factors of this kind and of the kinds of tests that measure them is largely available. Only by this kind of extension of intelligence testing can we do adequate justice to adult, human intellect.

Other extensions may also be very useful, for we are a long way from complete coverage of the intellectual factors in present tests. For differential prediction, and this includes the operation of vocational guidance, only single-factor scores will do complete justice in the description of individuals. As a necessary prelude to to the use of factor measures for such purposes, we need innumerable validation studies in which factors play an important role, studies such as those by Hills and others (**23**, **18**).

Summary

A listing of the factors that can be regarded as intellectual was made,

including those reported in French's summary of factors (**8**) appearing in 1951 and those reported since that time. Of approximately 40 such factors, seven are memory factors and the remaining ones have to do with thinking.

An attempt was made to formulate a system into which the factors seem to fall. The thinking factors were categorized under the general headings of cognition (discovery), production (convergent thinking and divergent thinking), and evaluation. The factors in each group can be arranged according to three kinds of content of thinking—figural, structural, and conceptual. In the cognition and production groups, a second principle of classification, cutting across the content principle, pertains to the kinds of things discovered or produced. In the memory group the second principle pertains to the kinds of things remembered—associations or substance. The result is a matrix of factors in each of the areas, with vacant cells. The vacancies suggest hypotheses for undiscovered factors.

In the general discussion, implications of the factors and their system were pointed out for factor theory and practice, for general psychological theory, and for the concept of intelligence and practices of intelligence testing.

REFERENCES

1. Adkins, Dorothy C., & Lyerly, S. B. Factor analysis of reasoning tests. Chapel Hill, N. C.: Dept. of Psychology, 1951.
2. Burt, C. The structure of mind: A review of the results of factor analysis. *Brit. J. Educ. Psychol.*, 1949, **19**, 100–111, 176–199.
3. Davis, P. C. A factor analysis of the Wechsler-Bellevue Intelligence scale, Form I, in a matrix with reference variables. Unpublished doctoral dissertation, Univer. of Southern California, 1952.
4. Ferguson, G. A. On learning and human ability. *Canad. J. Psychol.*, 1954, **8**, 95–112.
5. Fleishman, E. A. Dimensional analysis of psychomotor abilities. *J. exp. Psychol.*, 1954, **48**, 437–454.
6. Fleishman, E. A. & Hempel, W. E., Jr. Changes in factor structure of a complex psychomotor test as a function of practice. *Psychometrika*, 1954, **19**, 239–252.
7. Fleishman, E. A., & Hempel, W. E., Jr. The relation between abilities and improvement with practice in a visual discrimination reaction test. *J. exp. Psychol.*, 1955, **49**, 301–312.
8. French, J. W. The description of aptitude and achievement tests in terms of rotated factors. *Psychomet. Monogr.*, 1951, No. 5.
9. Green, R. F., Guilford, J. P., Christensen, P. R., & Comrey, A. L. A factor-analytic study of reasoning abilities. *Psychometrika*, 1953, **18**, 135–160.
10. Guilford, J. P. Progres recent dans la mesure des aptitudes. *Rev. Psychol. appl.*, 1952, **2**, 95–109.
11. Guilford, J. P. Some recent findings on thinking abilities and their implications. *Train. and Developm. Inform. Bull.*, 1953, **3**, 48–61.
12. Guilford, J. P. Factors in problem solving. *ARTC Instr. J.*, 1954, **4**, 197–204.
13. Guilford, J. P. Les dimensions de l'intellect. *Deliberations, Colloque internaternationale de l'analyse factorielle et ses applications.* Paris: Centre National Recherche Scientifique, 1955.
14. Guilford, J. P., Berger, R. M., & Christensen, P. R. A factor-analytic study of planning, I. Hypotheses and description of tests. *Rep. psychol. Lab.*, No. 10. Los Angeles: Univer. of Southern Calif., 1954.
15. Guilford, J. P., Berger, R. M., & Christensen, P. R. A factor-analytic study of planning, II. Administration of tests and analysis of results. *Rep. psychol. Lab.*, No. 12. Los Angeles: Univer. of Southern Calif., 1955.
16. Guilford, J. P., Christensen, P. R., Kettner, N. W., Green, R. F., & Hertzka, A. F. A factor-analytic study of Navy reasoning tests with the Air Force Aircrew Classification Battery. *Educ. psychol. Measmt*, 1954, **14**, 301–325.
17. Guilford, J. P., Kettner, N. W., &

CHRISTENSEN, P. R. A factor-analytic study across the domains of reasoning, creativity, and evaluation, I. Hypotheses and description of tests. *Rep. psychol. Lab.*, No. 11. Los Angeles: Univer. of Southern Calif., 1954.

18. GUILFORD, J. P., KETTNER, N. W., & CHRISTENSEN, P. R. The relation of certain thinking factors to training criteria in the U. S. Coast Guard Academy. *Rep. psychol. Lab.*, No. 13. Los Angeles: Univer. of Southern Calif., 1956.
19. GUILFORD, J. P., KETTNER, N. W., & CHRISTENSEN, P. R. A factor-analytic investigation of the factor called general reasoning. *Rep. psychol. Lab.*, No. 14. Los Angeles: Univer. of Southern Calif., 1955.
20. GUILFORD, J. P., KETTNER, N. W., & CHRISTENSEN, P. R. A factor-analytic study across the domains of reasoning, creativity, and evaluation, II. Administration of tests and analysis of results. *Rep. psychol. Lab.*, No. 16. Los Angeles: Univer. of Southern Calif., 1956.
21. GUILFORD, J. P., & LACEY, J. I. (Eds.) Printed classification tests. *AAF Aviat. Psychol. Program Res. Rep.* No. 5. Washington, D. C.: Government Printing Office, 1947.
22. HERTZKA, A. F., GUILFORD, J. P., CHRISTENSEN, P. R., & BERGER, R. M. A factor-analytic study of evaluative abilities. *Educ. psychol. Measmt*, 1954, **14**, 581–597.
23. HILLS, J. R. The relationships between certain factor-analyzed abilities and success in college mathematics. *Rep. Psychol. Lab.*, No. 15. Los Angeles: Univer. of Southern Calif., 1955.
24. JONES, L. V. A factor analysis of the Stanford Binet at four age levels. *Psychometrika*, 1949, **14**, 299–331.
25. JONES, L. V. Primary abilities in the Stanford Binet, age 13. *J. genet. Psychol.*, 1954, **84**, 125–147.
26. KARLIN, J. E. Music ability. *Psychometrika*, 1941, **6**, 61–65.
27. KELLEY, H. P. A factor analysis of memory ability. Princeton, N. J.: Educational Testing Service, 1954.
28. THURSTONE, L. L. A factorial study of perception. *Psychomet. Monogr.*, 1944, No. 4.
29. VERNON, P. E. The psychology of intelligence and G. *Bull. Brit. Psychol. Soc.*, 1955, No. 26.
30. WILSON, R. C., GUILFORD, J. P., & CHRISTENSEN, P. R. The measurement of individual differences in originality. *Psychol. Bull.*, 1953, **50**, 362–370.
31. WILSON, R. C., GUILFORD, J. P., CHRISTENSEN, P. R., & LEWIS, D. J. A factor-analytic study of creative-thinking abilities. *Psychometrika*, 1954, **19**, 297–311.

Received October 24, 1955

PSYCHOLOGICAL BULLETIN
Vol. 53, No. 4, 1956

PSYCHOTHERAPY AND THE PLACEBO EFFECT

DAVID ROSENTHAL
National Institute of Mental Health

AND JEROME D. FRANK
School of Medicine, Johns Hopkins University

It is by now generally recognized that all forms of psychotherapy yield successful results with some patients and that these successes depend to an undetermined extent on factors common to many types of relationship between patient and therapist. This poses a knotty problem for proponents of various specific forms of psychotherapy who are convinced that their successes result from their particular theory or technique and wish to convince others of this. As a result, problems of research design in psychotherapy have been receiving more and more critical attention in recent years, especially with reference to controls (**6**, **11**, **20**, **23**, **24**, **25**, **27**, **31**, **34**, **35**, **38**, **39**).

Certain general aspects of the psychotherapeutic relationship seem very similar to those responsible for the so-called placebo effect, which is well known to investigators of the therapeutic efficacy of medications. The purpose of this paper is to describe the placebo effect, discuss some of its implications for the evaluation of psychotherapy, and make some recommendations concerning research design in psychotherapy based on these considerations.

THE PLACEBO EFFECT

We have now participated in two separate investigations of the effectiveness of drugs on the symptomatic distress of psychiatric outpatients (**14**, **22**). Both studies involved the administration of a placebo, an inert agent outwardly indistinguishable from the agent being tested, as well as drugs. The physician never knew whether he was giving the patient drug or placebo. The patients were told that a new medicine had become available which, it was thought, might help them. The physicians rated symptoms on a 4-point scale of distress, with high reliability. In both studies a significant reduction of distress accompanied the taking of placebos, as shown in Table 1.

This phenomenon occurs with great regularity, not only with respect to the kinds of symptoms usually associated with psychologic illness, but with others as well. For example, in a study of vaccines for the common cold, there was found a reduction in the number of yearly colds of 55 per cent among those given vaccine and of 61 per cent among a control group who received injections of isotonic sodium chloride solution (**4**). Hillis (**15**) found placebos as effective as other agents in inhibiting the cough reflex. Wolf and Pinsky (**37**) studied medical outpatients suffering from peptic ulcer, migraine, muscle tension, headache, and tight muscles in the extremities. All were also tense and anxious. Twenty to thirty per cent felt better while taking placebos. Lasagna *et al.* (**19**) gave 1 ml. of saline by subcutaneous injection to surgical patients suffering from steady, severe wound pains and found that 30 to 40 per cent reported a satisfactory relief of pain. In a study by Jellinek (**18**) 60 per cent of 199 subjects with chronic headaches received relief from a placebo on one or more occasions.

TABLE 1

Symptom Distress Before Experiments and After a Trial on Placebos

Study	*N*	Drug tested	Mean distress scores		
			Before experiment	After placebo	Significance of difference
1st study	17	Mephenesin	25.58	15.88	.01
2nd study	16	Reserpine	34.06	24.69	.02

The placebo effect is not always favorable, but may also result in undesirable, distressful reactions. As far back as 1933, Diehl (**3**) using lactose placebos as a control for a variety of medications taken by mouth, found that some of his subjects receiving placebos developed nausea, faintness, and diarrhea. Sometimes this "toxic response" to placebos may even attain major proportions. Wolf and Pinsky (**37**) tell of one patient who had "overwhelming weakness, palpitation, and nausea within 15 minutes of taking her tablets." In another, "a diffuse itchy erythematous maculopapular rash developed after ten days of taking pills. A skin consultant considered the eruption to be typical dermatitis medicamentosa. After use of the pills was stopped, the eruption quickly cleared." A third patient developed epigastric pain followed by watery diarrhea, urticaria, and angioneurotic edema of the lips within ten minutes of taking her pills. One of our own patients, who had been tolerating a chronic syphilophobia fairly well, became acutely agitated shortly after placebo ingestion, bemoaning what the pills had done to him, and required hospitalization shortly thereafter.

Wolf and Pinsky (**37**) found that placebos produced more improvement in subjective than objective manifestations of anxiety and tension, but objective changes also occur. In our second study (**22**), 69 per cent of our patients showed decreased blood pressure and pulse readings following placebo, 19 per cent showed increased blood pressure, and 25 per cent showed a rise in pulse rate. Wolf (**36**) demonstrated clearly and convincingly that actual end-organ changes can follow placebo administration. This demonstration was made in a series of studies on the now-celebrated Tom, a human subject with a large gastric fistula, in whom it was possible to observe directly the gastric mucous membrane, correlating changes in color and turgidity with simultaneous measurements of gastric secretion and motor activity.

The placebo effect may actually reverse the normal pharmacologic action of a drug. For example, Wolf reports that Tom was repeatedly given Prostigmine, which induced abdominal cramps, diarrhea, as well as hyperaemia, hypersecretion, and hypermotility of the stomach. Subsequently, the same response occurred not only to tap water and lactose capsules, but also to atropine sulfate which usually has an *inhibiting* effect on gastric function. A pregnant patient with excessive vomiting showed the usual response of nausea and vomiting to ipecac. These manifestations were accompanied by cessation of normal gastric contractions. When ipecac was given through a tube with strong assurance that it would relieve her vomiting, gastric contractions were resumed at the same interval after ingestion of

the drug that they would normally have ceased, and her nausea and vomiting were relieved.

The placebo effect, in short, can be quite powerful. It can significantly modify the patient's physiological functioning, even to the extent of reversing the normal pharmacological action of drugs; and, as will be discussed below, it may be enduring. Placebo effects cannot be dismissed as superficial or transient. They often involve an increased sense of well-being in the patient and are manifested primarily by relief from the particular symptomatic distress for which the patient expects and receives treatment. Thus, the relief of any particular complaint by a given medication is not sufficient evidence for the specific effect of the medicine on this complaint unless it can be shown that the relief is not obtained as a placebo effect.

Implications of the Placebo Effect for Research in Psychotherapy

The giving of any medication may have certain meanings for a patient in terms of his relationship to his physician which may benefit his condition irrespective of the pharmacological action of the drug. For example, it may relieve the anxiety resulting from the distress caused by his illness (**10**). Wolf believes the effects of placebos on his patients "depended for their force on the conviction of the patient that this or that effect would result." The degree of the patient's conviction might be expected to be influenced by his previous experiences with doctors, his confidence in his physician, his suggestibility, the suggestibility-enhancing aspects of the situation in which the therapeutic agent is being administered, and his faith in or fear of the therapeutic agent itself. These attitudes are obviously relevant to psychotherapy.

Psychotherapists have theories of personality and psychotherapy and plan their therapeutic actions in the belief that these are the active agents which produce the desired results. Any favorable changes in patients consequent to a course of psychotherapy tend to be cited as evidence for the validity of the theory of personality and neurosis which underlie the rationale of the psychotherapy. In view of the above discussion it may well be that the efficacy of any particular set of therapeutic operations lies in their analogy to a placebo in that they enhance the therapist's and patient's conviction that something useful is being done. Patients entering psychotherapy have various degrees of belief in its efficacy, and this may be an important factor in the results of therapy, but this has not been studied, to our knowledge. We know that the authoritarian attitude of the physician can produce this conviction in some patients.

At first glance the attitudes found by Fiedler (**8**, **9**) to characterize experienced psychotherapists, viz. feelings of empathy for and closeness to the patient, an undemanding attitude, security, and the ability to "understand" the patient, seem diametrically opposed to the authoritarian attitude. It may be, however, that the therapeutic efficacy of these attitudes lies primarily in their ability to increase the confidence of certain patients in the ability of the therapist to help them. Lack of such confidence may be one of the reasons why patients of lower socioeconomic status fare less well in psychotherapy than patients higher in this scale (**16**, **29**), a talking therapy seeming to be beyond their comprehension and contrary to their conception of the doctor-patient relationship.

In this connection, the role of suggestion in psychotherapy has been emphasized for years, especially in therapies utilizing hypnosis, but suggestion effects have been thought by many since Freud to be superficial and transitory. We know of no experimental study which demonstrates that therapeutic effects based on insights or perceptual reorganization, which may also be suggested, are less superficial or less transitory.

It may be pointed out parenthetically that conviction of the helpfulness of therapy need not be equated with "motivation for therapy," which was investigated by Grummon (**13**) and Dymond (**5**) and found to have little relationship to success in psychotherapy. Patients are often sufficiently distressed to be strongly motivated to receive help, yet have little faith that a procedure such as psychotherapy can help them.

The similarity of the forces operating in psychotherapy and the placebo effect may account for the high consistency of improvement rates found with various therapies, from that conducted by physicians without psychiatric training to intensive psychoanalysis (**7**). This explanation gains plausibility from the fact that reported improvement rates for various series of neurotics treated by different forms of psychotherapy hover around 60 per cent (**1**). This is the same as that reported for the placebo effect in illnesses in which emotional components may play a major role such as "colds" (**3**) and headaches (**18**).

To show that a specific form of treatment produces more than a nonspecific placebo effect it must be shown that its effects are stronger, last longer, or are qualitatively different from those produced by the administration of placebos, or that it affects different types of patients. Our knowledge of all these matters is still fragmentary, but some beginnings have been made.

With respect to the strength and qualitative nature of the effects of therapy, one line of endeavor has been to study the physiological changes occurring during psychotherapy. Since physiological measures usually used to provide evidence of resistance or frustration (**26**, **33**) or similar psychological states during psychotherapy (**28**) may also be influenced by the placebo effect, one cannot conclude that demonstration of such physiological changes implies a greater *depth* of therapy or a more profound reorganization of the personality, unless we are willing to equate the placebo effect with such reorganization.

With respect to the duration of improvement, if it could be shown that the placebo effect is of shorter duration than changes specific to a given psychotherapy, this would provide one kind of evidence favoring that theory of psychotherapy. As far as we know, no study of the limits of duration of the placebo effect has been made. Our experiment with mephenesin vs. placebo covered four two-week periods. Figure 1 shows the curves for both agents for the eight weeks.

Figure 1 shows that the greatest decrease in distress following placebos was felt during the first two-week trial period. After that, a slight but statistically insignificant rise in distress occurred; and, at the end of eight weeks, the placebo effect was about as great as after two weeks. Unfortunately, our data yielded no information on how much longer it might have endured. If the effect is analogous to the relief of pain by placebos in patients with surgical wounds, we should expect it eventually to diminish. Lasagna *et al.* (**19**)

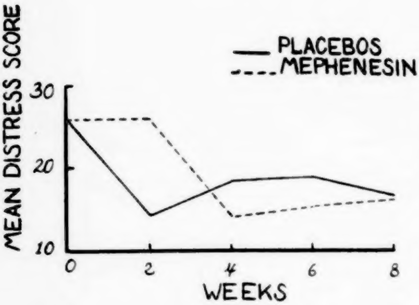

FIG. 1. EFFECTS OF MEPHENESIN AND PLACEBO ON SYMPTOMATIC DISTRESS OVER AN 8-WEEK PERIOD.

Total patients =17. At the 2-, 4-, 6-, and 8-week intervals, N for placebo =11, 6, 10, and 7 respectively, while N for mephenesin =6, 11, 7, and 10 respectively. For the 2- and 4-week periods, the dosage of mephenesin was 3 gms. per day; for the 6- and 8-week periods, 9 gms. per day.

found that as placebo therapy of such patients continued the relief experienced decreased.

Although the number of patients is too small to justify any conclusions, it is intriguing that the first dose of mephenesin seemed to counteract the placebo effect. In the study with reserpine (**22**), the only patients who failed to show a placebo effect were those who had received reserpine previously. It may be that any discomfort produced by a pharmacologically active agent tends to counteract the emotional state responsible for a placebo effect in susceptible patients. Analogously, an activity by the psychotherapist which disturbs the patient may conceivably counteract the placebo effect of psychotherapy with certain patients.

It would also be helpful to know if patients could be differentiated according to attributes which predisposed them to a positive or negative placebo effect. If patients who improved with a particular form of psychotherapy were all known to be positive placebo reactors, then the improvement could not be attributed to the specific form of treatment. If, however, they were known not to be positive placebo reactors, then any demonstrated improvement would constitute evidence of efficacy specific to the form of psychotherapy.

There is little known, however, with regard to the attributes of placebo reactors. Lasagna *et al.* (**19**) have made the first attempts to investigate this problem and report some attitudes and Rorschach categories which differentiated their reactors ($N=11$) from their nonreactors ($N=16$). However, only 14 per cent of their patients were consistent reactors, i.e., showed the effect with every placebo dose, and 31 per cent were consistent nonreactors, while 55 per cent showed the effect on some occasions but not on others. This contrasts with the findings of Jellinek (**18**) whose patients with headache were, for the most part, either in the always-relieved group or the never-relieved group, with only a small percentage of patients showing inconsistency of response. The apparent contradiction in findings may perhaps result from the difference in the cause of the pain in the two series or from other factors. In any case it indicates that the problem is a complex one needing much more study.

In the light of these considerations, any method of demonstrating the specificity of response to a given type of psychotherapy would have to provide an adequate control design. As far as we know, the study which has paid closest attention to the question of controls in research in psychotherapy is that of Rogers and his colleagues (**31**). They employed two different kinds of control groups. One was a group of nonclients who were simply given a battery of tests before and after specified time periods. The

other was a group of clients who were required to wait a specified period of time before beginning therapy. This group was tested at the beginning and end of the wait period, at the end of therapy, and after a follow-up period.

These procedures do not control for the placebo effect since neither control group was being subjected to any special procedures which could produce a reasonable expectancy in control subjects that certain changes should occur. The experimental group, however, could be expected to anticipate certain effects merely as a consequence of participating in the client-therapist interviews. Therefore, even though favorable changes could be demonstrated in their clients, the question of whether these were placebo effects could not be answered from such research design unless additional information were provided.

If we do not control for nonspecific factors like the placebo effect, we cannot know whether effects predicted from a theory lead to or result from improvement based on the nonspecific effect. Butler and Haigh (**2**), for example, report an increased correlation of perceived self with ideal self following client-centered therapy. The implicit inference is that the specific therapeutic method leads to this increased correlation which, in turn, contributes to amelioration of disability and distress.

It is conceivable, though, that as a result of a nonspecific placebo effect the client feels less disabled and distressed which, in turn, leads him to describe himself as more like his ideal self. Rogers' (**30**) findings of greater emotional maturity in successfully treated cases may be similarly explained, clients feeling less disabled and distressed due to a nonspecific placebo response and behaving consequently in ways which are less anxiety-determined and which are seen as more mature by others.

We would propose that the following conditions are optimal in planning research in psychotherapy:

1. A theory of personality and psychological distress (neurosis, maladjustment, etc.).
2. Predictions of effects in the patient or client consequent to psychotherapy, in accord with the theory.
3. Demonstration of a relationship between the predicted effects and some criterion of improvement.
4. Demonstration that the predicted effects and their relationship to the improvement criterion are not due primarily to the patient's conviction that therapy will help him. This will permit greater confidence that the relationship found is specific to the therapeutic technique derived from the theory.

Ideally, these conditions should obtain both for process and outcome research. There seems to be general agreement with regard to the first two conditions although Mackinnon (**21**) has some reservations about beginning with a theory rather than a hunch. Gordon *et al.* (**12**) have come to question the third condition, at least with respect to a "global" criterion of improvement.

The fourth condition has not been met in any research of which we are aware. It is not possible to set up an experiment precisely analogous to comparison of a medication with a placebo because there is no such thing as inert psychotherapy in the sense that placebos are pharmacologically inert. However, it may be possible to study the possible specific effects of any particular form of therapy by the use of a matched control group participating in an activity regarded as therapeutically inert from the standpoint of the theory of the therapy

being studied. That is, it would not be expected to produce the effects predicted by the theory. The "placebo psychotherapy" in this sense would be analogous to placebos in that it would be administered under circumstances and by persons such that the patients would expect to be helped by it.

Let us say that our theory is psychoanalytic and our predicted effect is an increased correlation between the moral values of the patient and the therapist (superego identification) and that we also expect an association between the increased correlation and a criterion of improvement (**32**). According to the theory, there is no reason to believe that control patients receiving, for example, relaxation therapy (**17**) will show the increased correlation of moral values with their therapist's moral values, nor should they show as much or as lasting improvement as the patients receiving psychoanalytic therapy of equal length. Such a design would constitute a fair test of the hypothesis based on the theory. In comparative studies where one type of psychotherapy is tested against another, differences found between them in predicted effects or amount, nature, and duration of improvement would not be explainable as placebo effects, if the condition could be met that patients had equal faith in the efficacy of the therapies and therapists to which they are assigned.

Summary and Conclusions

The literature on the therapeutic efficacy of drugs compared with placebos is briefly reviewed, and its relevance for research in psychotherapy considered. It is concluded that improvement under a special form of psychotherapy cannot be taken as evidence for: (*a*) correctness of the theory on which it is based; or (*b*) efficacy of the specific technique used, unless improvement can be shown to be greater than or qualitatively different from that produced by the patients' faith in the efficacy of the therapist and his technique—"the placebo effect." This effect may be thought of as a nonspecific form of psychotherapy and it may be quite powerful in that it may produce end-organ changes and relief from distress of considerable duration.

To show that a specific form of psychotherapy based on a theory of personality and neurosis produces results not attributable to the nonspecific placebo effect it is not sufficient to compare its results with changes in patients receiving no treatment. The only adequate control would be another form of therapy in which patients had equal faith, so that the placebo effect operated equally in both, but which would not be expected by the theory of therapy being studied to produce the same effects. We need to learn more about the nature of the placebo effect, the conditions giving rise to it, and the attributes of patients most susceptible or resistant to it so that we may obtain a better understanding of the role of nonspecific factors in psychotherapy.

References

1. Appel, K. E., Lhamon, W. T., Myers, J. M., & Harvey, W. A. Long term psychotherapy. In *Psychiatric treatment. Proc. Ass. Res. Nerv. Ment. Dis.* Dec. 14, 15, 1951, New York. Baltimore: Williams & Wilkins, 1953, 21–34.
2. Butler, J. M., & Haigh, G. V. Changes in the relation between self-concepts and ideal concepts consequent upon client-centered counseling. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954. Pp. 55–75.

3. Diehl, H. S. Medical treatment of the common cold. *J. Amer. Med. Ass.*, 1933, **101**, 2042–2045.
4. Diehl, H. S., Baker, A. B., & Cowan, D. W. Cold vaccines, further evaluation. *J. Amer. Med. Ass.*, 1940, **115**, 593–594.
5. Dymond, Rosalind F. Adjustment changes in the absence of psychotherapy. *J. consult. Psychol.*, 1955, **19**, 103–107.
6. Edwards, A. L., & Cronbach, L. J. Experimental design for research in psychotherapy. *J. clin. Psychol.*, 1952, **8**, 51–59.
7. Eysenck, H. J. The effects of psychotherapy—An evaluation. *J. consult. Psychol.*, 1952, **16**, 319–324.
8. Fiedler, F. E. The concept of the ideal therapeutic relationship. *J. consult. Psychol.*, 1950, **14**, 239–245.
9. Fiedler, F. E. A comparison of therapeutic relationships in psychoanalytic, nondirective, and Adlerian therapy. *J. consult. Psychol.*, 1950, **14**, 436–445.
10. Frank, J. D. Psychotherapeutic aspects of symptomatic treatment. *Amer. J. Psychiat.*, 1946, **103**, 21–25.
11. Greenhill, M. H., Ford, L. S., Olson, W. C., Ryan, W. C., Whitman, S., & Skeels, H. M. *Evaluation in mental health*, Bethesda: National Institute of Mental Health, 1955.
12. Gordon, T., Grummon, D. L., Rogers, C. R., & Seeman, J. Developing a program of research in psychotherapy. In C. R. Rogers & Rosalind F. Dymond (Eds.) *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954. Pp. 12–34.
13. Grummon, D. L. Personality changes as a function of time in persons motivated for therapy. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954, 238–255.
14. Hampson, J. L., Rosenthal, D., & Frank, J. D. A comparative study of the effects of mephenesin and placebo on the symptomatology of a mixed group of psychiatric outpatients. *Bull. Johns Hopkins Hosp.*, 1954, **95**, 170–177.
15. Hillis, B. R. The assessment of cough-suppressing drugs. *Lancet*, 1952, **1**, 1230–1232.
16. Imber, S. D., Nash, E. H., & Stone, A. R. Social class and duration of psychotherapy. *J. clin. Psychol.*, in press.
17. Jacobson, E. *Progressive relaxation.* Chicago: Univer. of Chicago Press, 1938.
18. Jellinek, E. M. Clinical tests on comparative effectiveness of analgesic drugs. *Biometrics Bull.*, 1946, **2**, 87.
19. Lasagna, L., Mosteller, F., Felsinger, J. M., & Beecher, H. K. A study of the placebo response. *Amer. J. Med.*, 1954, **16**, 770–779.
20. Lebo, D. The present status of research on nondirective play therapy. *J. consult. Psychol.*, 1953, **17**, 177–183.
21. Mackinnon, D. W. Fact and fancy in personality research. *Amer. Psychol.* 1953, **8**, 138–145.
22. Meath, J. A., Feldberg, T. M., Rosenthal, D., & Frank, J. D. A comparative study of reserpine and placebo in the treatment of psychiatric outpatients. (Unpublished manuscript.)
23. Morse, P. W. A proposed technique for the evaluation of psychotherapy. *Amer. J. Orthopsychiat.*, 1953, **4**, 716–731.
24. Mosak, H. H. Problems in the definition and measurement of success in psychotherapy. L. W. Wolff & J. A. Precker: *Success in psychotherapy.* New York: Grune & Stratton, 1952, 1–25.
25. Mowrer, O. H. (Ed.). *Psychotherapy: theory and research.* New York: Ronald, 1953.
26. Mowrer, O. H., Light, B. H., Luriz, Z., & Zeleny, M. P. Tension changes during psychotherapy, with special reference to resistance. In O. H. Mowrer (Ed.) *Psychotherapy: theory and research.* New York: Ronald, 1953. Pp. 546–640.
27. Oberndorf, C. P., Greenacre, Phyllis, & Kubie, L. Symposium on the evaluation of therapeutic results. *Int. J. Psychoanal.*, 1948, **29**, 7–33.
28. O'Kelly, L. I. Physiological changes during psychotherapy. In O. H. Mowrer (Ed.) *Psychotherapy: theory and research.* New York: Ronald 1953, 641–656.
29. Redlich, F. C., Hollingshead, A. B., Roberts, B. H., Robinson, H. A., Freedman, L. Z., & Myers, J. K. Social structure and psychiatric disorders. *Amer. J. Psychiat.*, 1953, **109**, 729–734.
30. Rogers, C. R. Changes in the maturity of behavior as related to therapy. In C. R. Rogers & Rosalind F. Dymond (Eds.) *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954. Pp. 215–237.
31. Rogers, C. R., & Dymond, Rosalind F.

(Eds.) *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954.

32. Rosenthal, D. Changes in some moral values following psychotherapy. *J. consult. Psychol.*, 1955, **19**, 431–436.
33. Thetford, W. N. An objective measurement of frustration tolerance in evaluating psychotherapy. In W. Wolff & J. A. Precker, *Success in psychotherapy.* New York: Grune & Stratton, 1952, 26–62.
34. Thorne, F. C. Rules of evidence in the evaluation of the effects of psychotherapy. *J. clin. Psychol.*, 1952, **8**, 38–41.
35. Watson, R. I., Mensch, I., & Gildea, E. F. The evaluation of the effects of psychotherapy III. Research design. *J. Psychol.*, 1951, **32**, 293–308.
36. Wolf, S. Effects of suggestion and conditioning on the action of chemical agents in human subjects—the pharmacology of placebos. *J. clin. Invest.*, 1950, **29**, 100–109.
37. Wolf, S., & Pinsky, R. H. Effects of placebo administration and occurrence of toxic reactions. *J. Amer. Med. Ass.*, 1954, **155**, 339–341.
38. Wolff, W., & Precker, J. A. *Success in psychotherapy.* New York: Grune & Stratton, 1952.
39. Zubin, J. Design for the evaluation of therapy. In *Psychiatric treatment. Proc. Ass. Res. Nerv. Ment. Dis.* Dec. 14, 15, 1951, New York. Baltimore: Williams & Wilkins, 1953. Pp. 10–15.

Received July 22, 1955

Psychological Bulletin
Vol. 53, No. 4, 1956

DRIVE THEORY AND MANIFEST ANXIETY

JANET A. TAYLOR
Northwestern University

In recent years a number of studies involving human *S*s have been devoted to testing the implications of certain Hullian notions concerning the relationship between performance in learning situations and level of total effective drive (*D*). In these investigations drive has been defined in terms of scores on a manifest anxiety scale (**41**). In view of the growing experimental literature concerning these hypotheses since their initial statement by Taylor (**40**) and Taylor and Spence (**43**), an attempt to outline the theory as it is presently conceived by the Iowa group and to evaluate the evidence concerning it seems to be in order.

Before proceeding with these matters, however, certain misunderstandings which have arisen concerning the purpose of this work should be mentioned. First, although groups have been selected exclusively on the basis of scores on the Manifest Anxiety Scale (hereafter designated as MAS) the interest of the Iowa group has *not* been in investigating anxiety as a phenomenon, but rather in the role of drive in certain learning situations. The assumption has been made that anxiety scores are related in some manner to drive level, but in terms of the major theoretical interests of this group, any other acceptable specification of drive (eg., hunger) could be used in experimental tests of the hypotheses about the effect of drive level. Further, as Farber (**6**) has pointed out, no attempt has ever been made to claim that the only difference between individuals receiving different scores on the MAS is in drive level or that all performance differences could be explained by drive. Undoubtedly there are many characteristics other than drive level on which anxious and nonanxious *S*s differ; the investigation of these additional properties of anxiety groups and their influence on performance is certainly both legitimate and important, but it simply has not been the interest of the proponents of the drive theory.

A second point that should be clarified has to do with the MAS. The construction of the test was not aimed at developing a clinically useful test which would diagnose anxiety, but rather was designed solely to select *S*s differing in general drive level. Thus the question of the scale's "validity" (i.e., its agreement with clinical judgments) is in a sense irrelevant to the experimental purposes for which the test was developed. In light of this, the test might better have been given a more noncommittal label, such as a measure of emotionality, although the fact that the items on the scale were selected by clinicians as referring to manifest anxiety as it is described psychiatrically does not make the title completely inappropriate nor a relationship between clinical judgments and MAS scores unexpected. Certainly the generality of the experimental findings with the MAS would be increased if correlations were found with other definitions and such attempts will be discussed in a later section. However, regardless of the results of such studies, it should be clearly understood that "manifest anxiety" has been de-

fined operationally only in terms of test scores and will be so employed, unless otherwise indicated, in the present paper.

Drive Theory

As stated earlier, the purpose of the Iowa group has been to investigate the effects of varying drive level on performance in learning situations. Actual experimentation has involved two independent problems: (*a*) specification of the conditions under which drive differences are said to appear, and (*b*) the theory concerning the effects of drive level on behavior once drive has been aroused. The first problem concerns the postulated relationship between the MAS and drive level, the second between drive (or anxiety) level and performance in various situations. Since the two are separate matters, an outline of the theory concerning the influence of drive will be given first and the hypothesized relationship between drive and MAS scores considered at a later point.

According to Hull (**15**), all habits (H) activated in a given situation combine multiplicatively with the total effective drive state (D) operating at the moment to form excitatory potential $E[E = f(H \times D)]$. Total effective drive, in the Hullian system, is determined by the summation of all extant need states, primary and secondary, irrespective of their source and their relevancy to the type of reinforcement employed. Since response strength is determined in part by E, the implication of varying drive level in any situation in which a single habit is evoked is clear: the higher the drive, the greater the value of E and hence of response strength. Thus in simple noncompetitional experimental arrangements involving only a single habit tendency the performance level of high-drive *S*s should be greater than that for low-drive groups.

Higher drive levels should not, however, always lead to superior performance (i.e., greater probability of the appearance of the correct response). In situations in which a number of competing response tendencies are evoked, only one of which is correct, the relative performance of high and low drive groups will depend upon the number and comparative strengths of the various response tendencies. Predictions concerning the performance of the groups in such complex tasks involve the introduction of additional Hullian concepts: oscillatory inhibition (O) and threshold (L).

The concept of O was introduced by Hull (**15**) in an attempt to allow for statement, within his system, of the intra-individual variability in behavior that occurs, presumably, because of uncontrolled variations from instant to instant within the organism and in his environment. The value of O is said to vary from moment to moment, the distribution of O values for a group of (like) individuals on any trial forming a normal probability function. O is further assumed to play an inhibitory role, its value being subtracted from excitatory potential (E), thus yielding momentary excitatory potential ($\dot{E}$). In order for $\dot{E}$ to activate a response, it must attain a minimum or threshold value (L), a value that is presumably the same for all similar habit tendencies evoked in a given situation. Thus $R = f(\dot{E}) = f(E\text{-}O\text{-}L)$.

In any task in which a stimulus tends to evoke a number of competing responses the response that will appear on a given occasion will be the one with the highest suprathreshold momentary excitatory strength ($\dot{E}$)

at that moment. Other things being equal, of course, the response with the greatest *H* and hence *E* value will have a greater probability of occurring than any other response.

Adding the notion of differing drive level to this conception, we see that the probability of appearance of the correct response involves an interaction between drive level and the number and comparative strengths of the correct and incorrect tendencies. When the correct response is weaker (i.e., has less *H*) than one or more of the competing response tendencies, high-drive groups should be inferior in performance to low-drive *S*s. That is, because of the multiplicative relationship between habit strength and drive, the stronger incorrect tendencies gain relatively more *E* than the correct tendency in the case of high drive *S*s than in low drive, thus leading to a greater probability of occurrence of one of the stronger incorrect responses in the high-drive group. Further, the possibility exists that under a high-drive level new competing responses with very weak habit strengths may be brought over the threshold value of *E* with the consequence that the probability of occurrence of the correct response is lowered relative to that in a low-drive condition.

At the other extreme, the correct response tendency may be highest in the hierarchy and relatively strong when compared to the incorrect. In such a situation, which is comparable to the case in which but a single habit is aroused, the *E* value for the correct response would be relatively greater than the other responses in the hierarchy for the high-drive group than for the low-drive, leading to the prediction of the superiority of performance of such subjects.

It should be obvious, then, that maximum inferiority of high-drive *S*s would be expected when a large number of competing tendencies are present and the correct tendency is both relatively weak and low in the hierarchy. As the strength of the correct tendency increases relative to the incorrect, high-drive groups should become less inferior and eventually superior in performance to low-drive groups. The exact point of equality would be difficult to specify. Even when the correct response is highest (though not strongly dominant) in the hierarchy, high-drive *S*s could still conceivably be inferior in some instances since a greater number of suprathreshold tendencies could more than offset the advantage of the relatively higher *E* value of the correct response for these individuals.[1]

An important consideration that should be noted about making predictions concerning the effect of drive level upon performance in actual experimental situations is that a behavioral analysis of the situation must have been made; only in experimental arrangements in which the results, independent of drive level, permit statements in terms of competing S-R tendencies are deductions from the theory possible. While the majority of investigations designed to

[1] In a recent review Child (**3**) incorrectly interpreted the theoretical analysis outlined above as involving the sudden introduction of *O* and *L* for the situation in which the correct response is highest in the hierarchy. These concepts are of course assumed to be operating in all situations, including the noncompetitional one in which but a single response tendency is being evoked. No appeal was made to these constructs in the latter instance, however, since their inclusion would not affect the predictions. Mention might also be made of other constructs in the Hullian system (e.g., *I*, *V*, *K*, etc.): it has been assumed that these are of equal value for all drive groups and that a consideration of their values would not result in changing any prediction.

test implications of these derivations concerning drive level have utilized tasks for which analyses in S-R terms had already been made and found to be useful, occasionally an experiment appears in which the investigator attempts to evaluate the total theory by comparing groups on a task which is poorly understood (and for which little or no rationale is presented) or which clearly involves the introduction of variables not included in the theory. The accumulation of empirical evidence concerning the performance of different groups in any situation or attempts to incorporate additional variables within any theoretical framework are certainly to be encouraged, but statements that the results of such studies refute or confirm theoretical expectations are unwarranted when there is no evidence that the boundary conditions imposed by the theory are met.

Drive and Anxiety

The use of the MAS to select groups that are postulated to differ in drive level in an experimental situation has rested on the assumption that scores on the scale are in some manner related to emotional responsiveness, which, in turn, contributes to drive level. Two alternative hypotheses have been entertained concerning the conditions under which emotionality is evoked. One is that test scores reflect differences in a chronic emotional state so that individuals scoring high on the scale tend to bring a higher level of emotionality or anxiety "in the door" with them than do *S*s scoring at lower levels (**40**). A second alternative conception is that MAS scores reflect different potentialities for anxiety arousal, high scoring *S*s being those who tend to react more emotionally and adapt less readily to novel or threatening situations than do low scorers (**28**, **37**). According to the first hypothesis differences among anxious and nonanxious groups (providing other conditions imposed by the theory are met) should be found whether or not there is any "threat," in the form of noxious stimulation, fear of failure or the like, in the situation. Thus, for example, the performance of anxious *S*s should be superior to the nonanxious in both classical defense conditioning, in which a noxious stimulus is employed, and in reward conditioning into which no objective threat has been introduced. In the case of the second conception, differences would be expected in the performance of anxiety groups only in those situations in which some threat is present. Should this be the correct conception, exact specification of the conditions thought to be sufficient to evoke anxiety would be necessary in order to test hypotheses concerning the role of drive. Available evidence suggests that the magnitude of differences among groups may be related to the level of noxious stimulation employed (**37**), or to stress-producing instructions (**10**,**19**), suggesting that differences in drive level among groups may depend at least in part upon situational factors. However, the picture is complicated by the results of a number of studies in which differences among anxiety groups have been found in the absence of noxious stimulation or instructions designed to produce stress (**8**, **24**, **25**, **26**, **42**).

Most investigators have not explicitly considered this issue, assuming either that anxiety scores reflect a chronic level of emotionality or that factors are present in the typical laboratory experiment that result in different anxiety levels among groups. For purposes of evaluating those stu-

dies in which degree of stress has not been under investigation, the assumption will tentatively be made here that in all situations, individuals scoring high and low on the anxiety scale will differ in drive level, for whatever reason. The evidence more directly concerned with the conditions of anxiety-arousal will be considered at a later point.

Experimental Evidence

Classical conditioning. Classical conditioning is said to be a noncompetitional situation in which but a single response tendency is being acquired; theoretical expectation therefore is that anxious groups will perform at a higher level than nonanxious. The results of a number of studies of eyelid conditioning using groups with extreme scores on the MAS[2] have upheld these predictions, anxious *S*s showing a greater number of CR's than nonanxious (**11**, **35**, **37**, **38**, **39**, **40**). In all cases but one (**11**), these differences were statistically significant, the exception involving the use of only 10 *S*s per group, considerably fewer than were employed in other investigations. Data from eyelid conditioning studies performed in the Iowa laboratories and elsewhere (**39**) are also available from *S*s scoring throughout the entire range of anxiety scores rather than only at the two extremes. The relationship between anxiety and conditioning scores has been uniformly found to be monotonic although not always linear, middle-anxiety *S*s tending to show a performance level closer to the low-scoring than the high-scoring groups. The magnitudes of the correlation coefficients obtained have been in the neighborhood of .25, thus indicating that relatively little of the variance among *S*s can be accounted for in terms of anxiety scores. In view of the low correlation and the monotonic relationship between the two variables, continued use of extreme groups only for research purposes in such situations seems justified.

A conditioning study employing a response other than the eyeblink has also been reported in the literature. An investigation by Bitterman and Holtzman (**1**) utilized the PGR technique which, like the eyelid situation it will be noted, involves defense conditioning. After dividing a group of randomly selected college students into the upper and lower 50% on the basis of MAS scores, these investigators found a slight but statistically insignificant superiority in conditioning level on the part of their anxious *S*s. Since their anxious group included individuals with scores considerably lower than those in the investigations referred to above, this lack of statistical significance is not too surprising.

Several studies are available concerning differential conditioning, also in the eyelid situation (**11**, **34**, **36**). The predictions derived from the theory in this instance are that anxious *S*s should exhibit a greater excitatory strength both to the positive (reinforced) CS and to the negative (nonreinforced) CS and further, that the difference in excitatory strengths of the two stimuli should be greater for the anxious group. By transforming all raw data into excitatory strength values, Spence and his colleagues (**34**, **36**) have attempted to test these predictions in some five

[2] In almost all of the studies involving the MAS, a comparison has been made of extreme scorers, typically the 20th percentile or below (nonanxious) and 80th percentile or above (anxious) in terms of a standardization group of college students (**41**). Use of the terms "anxious" and "nonanxious" groups here should be understood to refer to such extremes unless otherwise indicated.

separate instances. In each case, the excitatory strength to the positive *CS* during differential conditioning was significantly greater for anxious *S*s, as was expected. The results concerning the remaining two predictions were not so clear-cut. In four out of five independent instances the excitatory strength to the negative stimulus was greater for the anxious *S*s but in no case was the difference significant. In all five cases the difference between excitatory strengths was in the expected direction but was significant in only one instance. While the results of these studies tend to lend some support to the theory, somewhat contradictory findings have been reported by Hilgard, Jones, and Kaplan (**11**). As mentioned earlier, contrary to other studies of simple eyelid conditioning, these investigators found only a slight, statistically insignificant superiority for anxious *S*s during training to the positive CS. During differential conditioning, the anxious group continued to exhibit an insignificant superiority to the nonanxious on the positive CS. However, the responses of the anxious *S*s to the negative CS were significantly greater as would be expected by drive theory.

Stimulus generalization. Stimulus generalization, to which differential conditioning is related, has been investigated more directly by Rosenbaum (**28**) and Wenar (**48**). Rosenbaum found greater responsiveness to generalized stimuli in a spatial situation for an anxious group than for a nonanxious group, as would be predicted by drive theory, but only in the case of *S*s given strong intermittent shock during their performance; for groups given a weak shock or buzzer, no significant differences emerged. After training groups of anxious and nonanxious *S*s on a key-pressing response to a strong shock, weak shock or a buzzer presented at regular intervals, Wenar (**48**) measured the reaction time to these stimuli in a test series in which the intervals of presentation were longer or shorter (temporal generalization) than those employed during training. Reaction time was related significantly to both stimulus intensity and anxiety level, response time being quicker as these variables increased.

Maze learning. The first study to be concerned with demonstrating that the relative performance of anxious and nonanxious *S*s is a function of degree of interference within a task was reported by Taylor and Spence (**43**), who used a type of serial verbal maze. On the assumption that errors in such a situation are largely the result of interfering response tendencies, due to remote associations, etc., it was expected that anxious *S*s would make more errors and take more trials to reach a criterion than nonanxious. The results of this study and of a subsequent investigation by Farber and Spence (**8**) with a stylus maze have confirmed these hypotheses, the greater number of errors and trials to criterion being made by the anxious groups. An additional prediction was also made for these maze data, namely that the degree of inferiority of the anxious *S*s in comparison to the nonanxious should be positively related to difficulty of the choice point. In both studies, significant rank-order correlations were obtained between the difference in number of errors between groups on an individual choice point and the difficulty of that point. Although these results tend to confirm theoretical expectation, some discrepancy between prediction and the experimental findings occurred on the easiest choice points. In each

investigation, the small number of errors on the easiest two or three points suggests the presence of few interfering tendencies so that the anxious might be expected to be superior in performance. Even here, however, they tended to be inferior.

In addition to the two studies utilizing extreme groups, one study of stylus maze learning involving the entire range of anxiety scores has been reported. After splitting a randomly selected group of college students into 7 anxiety groups according to their MAS scores, Matarazzo *et al.* (**24**) found a linear relationship ($r=.25$) between anxiety and trials to the criterion on the maze.

While the investigations reported above have found differences between anxiety groups on maze performance, Hughes, Sprague, and Bendig (**14**), utilizing extreme groups, failed to duplicate these results with several serial verbal mazes. Different from the Taylor and Spence study in which the typical 2-second rate of stimulus presentation was employed, Hughes *et al.* used a 4-second rate in all cases. Previous investigations have demonstrated (**12**) that performance is positively related to the interstimulus interval in serial learning but since the effects of this variable are poorly understood, the implications of the failure to find differences between anxiety groups with the 4-second condition are not clear. One possibility, based on the assumption that differences in anxiety level are largely determined by situational factors, is that under longer time intervals, stress upon *S*s, and hence upon differences in emotionality between anxious and nonanxious, is minimized.

Verbal learning. Rather than attempting to demonstrate an interaction between anxiety level and degree of interference by examining individual items within a single task, as was done in the maze studies, Montague (**25**) formed three different lists of serial nonsense syllables which, because of varying degrees of formal intralist similarity and association value of the syllables, presumably differed in the amount of intralist interference. A significant interaction was found between anxiety and list, an anxious group being significantly superior in performance to nonanxious on the list for which similarity was low and association value high, and the position being reversed for groups given a list of high similarity and low association value. Similar findings have been reported by Lucas (**19**) in a study in which *S*s were asked to recall lists of consonants read to them. As the number of duplicated consonants within a list was increased, anxious *S*s showed a significant decrease in the amount recalled while the performance of the nonanxious was not affected.

While a number of investigators have employed serial learning tasks, from the point of view of testing the implications of drive theory, the paired-associate technique seems to be preferable. Whereas intralist interferences due to such factors as remote associations are inherently part of serial learning and are thus difficult to manipulate, the use of discrete S-R pairs permits more precise control of the number and strength of the response tendencies elicited by each stimulus. Turning to the investigations that have employed this paired-associate arrangement, several studies have attempted to minimize the presence of competing response tendencies and thus to demonstrate the performance superiority of anxious *S*s. In one, Taylor and Chapman (**42**) chose nonsense syllables with

low formal similarity, in an attempt to provide a noncompetitional arrangement in which each stimulus tended to evoke only its own response. As expected, on two lists for which such low similarity obtained, anxious *S*s were significantly superior in performance to nonanxious. Similar superiority of anxious *S*s has been reported by Spence (**33**) on an adjective list in which the association between each S-R pair was presumed to be initially strong and minimum similarity existed among pairs. In a second part of this investigation, an attempt was made to maximize the number of competing tendencies by having a high degree of synonymity among stimuli. As predicted, an anxious group in this case was inferior.

The initial strength of association between S-R was also manipulated by Ramond (**26**) in an investigation involving a variation of the standard paired-associate technique. Each stimulus, an adjective, had connected with it two response words, one judged to be highly associated with the stimulus and the other with no discernible association. Each type of response was correct for half of the items. When the low association responses were correct, anxious *S*s were expected to perform at a lower level than nonanxious because of the greater interference of the strong, incorrect response for this group. The results confirmed this prediction. Theoretical expectations for the situation in which the stronger response was correct are not so clear-cut since the arrangement of the list made it likely that as learning took place the low association responses would interfere occasionally with the high association response because of stimulus generalization. Thus, while anxious *S*s might be expected to be superior early in learning, they might lose this superiority as the weak responses are learned and provide competition. The results lent some support to these expectations, anxious *S* first being superior and then inferior to nonanxious although the over-all difference between groups did not reach statistical significance.

Anxiety Scores and Their Relationship to Stress

As was indicated earlier, two alternative hypotheses have been entertained concerning the difference between *S*s scoring high and low on the MAS with respect to anxiety: that such groups have different levels of chronic anxiety or that the groups instead differ in their emotional reactiveness to anxiety-evoking stimuli present in a situation.

The studies of verbal learning just discussed indicate that whether due to chronic or situational factors, differences between high and low scoring *S*s cannot be said to be produced only when stress is deliberately introduced into the situation, either by means of noxious stimulation as in the case of defense conditioning or by the administration of stress-provoking instructions (e.g., reports of failure). Consideration of the studies into which some threatening stimulation has been introduced may, however, throw some light onto the question as to whether differences in anxiety among groups could depend, at least in part, on situational variables.

Should situational factors play a role in determining differences in emotionality among anxiety groups, the strength of the UCS in classical conditioning might be expected to be related to such group differences. A comparison of three experiments of eyelid conditioning from the Iowa laboratory involving a relatively strong, medium, and mild UCS, re-

spectively, was made by Spence and Farber (**35**). Examination of the mean conditioning scores reveals that while intensity of the UCS tended to be related to performance, the magnitude of the difference between anxious and nonanxious remained relatively constant under the different intensities. Different results were obtained by Spence and his associates (**37**) in a study specifically undertaken to evaluate the effect of the strength of noxious stimulation on anxiety groups. In this investigation the *S*s, selected without reference to their anxiety scores, were conditioned with a relatively weak UCS, but one group was given occasional electric shocks between trials, another threatened with shock, and a third trained under neutral conditions. These latter *S*s, run under neutral conditions, gave fewer CR's than the other groups, especially in earlier trials. When *S*s were later divided into the upper and lower 50 per cent according to anxiety scores, it was found that while the high-scoring group conditioned without shock or threat of shock exhibited only a slight, statistically insignificant superiority in conditioning performance, the difference between anxiety groups was highly significant for *S*s with whom shock or threat of shock was employed.

The previously mentioned studies of stimulus generalization by Rosenbaum (**28**) and Wenar (**48**) were also concerned with variations in the intensity of noxious stimulation, in both cases a buzzer and two intensities of shock being employed. While Rosenbaum found a significant difference between groups only when strong shock was used, Wenar's results (with a somewhat different experimental arrangement) indicated a greater responsiveness for the anxious group under all three conditions. Furthermore, the magnitude of the difference between groups was unaffected by stimulus intensity.

Turning to verbal learning, Deese, Lazarus, and Keenan (**4**) have reported a study in which the effect of electric shock on serial learning was investigated. Here it was found that nonanxious groups given intermittent shocks performed at a significantly lower level than a nonanxious control group run under neutral conditions. In contrast, the performance of the anxious groups remained relatively constant, *S*s run under shock not differing from their control group. Further, when all conditions were combined, the performance of the anxious was significantly superior to the nonanxious.[3] Thus, while the differences between groups increased under shock, they were due to the disruptive effect of the shock on the ***non-anxious*** *S*s.

Quite in contrast to the results of Deese *et al.* are the findings obtained

[3] Although, presumably, the serial list was of relatively low intralist similarity, it is difficult to tell from the writers' description what drive theory would have predicted concerning the performance of the anxiety groups, independent of the stress factor. In a second, parallel, experiment involving a more difficult list (12 consonant syllables composed of only 5 consonants) presented for a standard 12 trials, Lazarus, Deese, and Hamilton (**17**) found no differences among groups either as a function of anxiety scores or of shock-no-shock conditions. While these results appear superficially to be contradictory both to drive theory (which would expect inferiority of anxious *S*s) and to the results of the first study with respect to the influence of shock, inspection of their data indicates that all groups averaged only about one correct response per trial. Since so little learning took place it is not surprising to have no differences in performance among groups. For this reason it is felt that the study does not provide very meaningful evidence on the effects of either anxiety level or shock on task performance.

by Gordon and Berlyne (**10**) in an investigation of verbal learning utilizing psychological stress rather than noxious stimulation. After being told that the tasks were measures of intelligence and that their performance on a paired-associate list was above average, anxious and nonanxious groups did not differ significantly in amount of negative transfer on a second paired-associate list. An anxious group told that their first list performance was below average, however, exhibited significantly more negative transfer than did a comparable nonanxious group. Finally, in the Lucas study (**19**) mentioned earlier in which the recall of consonants lists varying in number of duplications was investigated, the effects of varying numbers of reports of failure to meet expected standards were also studied. While nonanxious *S*s increased the amount recalled with greater numbers of failure experiences, the anxious groups did significantly worse.

As may be seen, the available evidence does not present a clear-cut picture with respect to the effects of stress. Summarizing first those investigations involving noxious stimulation, the results indicate that with one exception (**4**) the performance of all *S*s tends to be affected in the same direction as is found with an increase in anxiety (MAS) level. The magnitude of the difference between anxious and nonanxious *S*s either remains constant with greater degrees of stimulation or is increased. The data from the two studies employing psychological stress (in both cases defined by telling *S* he had failed to achieve adequate standards on an intelligence test) have revealed somewhat different relationships. In both instances (**10**, **19**) the performance of anxious *S*s under stress was significantly worse than the anxious group tested under neutral conditions while the performance of nonanxious *S*s was in one case the same and in the second better than the control group. Thus, the magnitude of the difference between anxiety groups was greater under stress than under neutral conditions.

The available evidence suggests then that situational sources of stress may play a role in determining the difference in anxiety level between *S*s scoring at the extremes of the MAS. Whether the differences between groups in the verbal learning studies into which no objective stress had been introduced by the experimenter reflect chronic anxiety level or unidentified sources of threat remains an open question. Speculating on this point, to many college sophomores psychology experiments per se may be seen as somewhat threatening, particularly when the task could be interpreted as reflecting on their personality or intelligence. It is perfectly possible that in experimental arrangements involving no noxious stimulation or stress-inducing instructions which call upon skills not particularly valued by college students, differences between groups might disappear.[4]

Using the results of these studies involving stress to attempt to determine the source of anxiety differences between high- and low-scoring *S*s or, for that matter, to test drive theory, involves the assumption that the only effect of stress in any situation is to increase drive level or, at least, that

[4] A study of classical reward conditioning of the salivary response by Bindra, Paterson, and Strzelecki (On the relation between anxiety and conditioning, *Canad. J. Psychol.*, 1955, **9**, 1–6) which appeared after this review was written confirms this suggestion. No difference was found between anxious and nonanxious groups.

anxious and nonanxious groups do not respond differentially to stress except with respect to anxiety or drive. Although no systematic exploration has been made of the relationship between degree of noxious stimulation and performance on various types of tasks, an examination of the general literature concerning the effect of such stimulation in nonverbal, noncompetition situations lends some credibility to this assumption (**32**). It is important to note that with one exception (**4**) the studies of the effects of noxious stimuli on anxious and nonanxious *S*s have employed tasks of this type.

In contrast, the literature concerning studies of psychological stress (e.g., ego-involving instructions, reports of failure), most of which have employed quite complex tasks, suggests that factors other than or in addition to drive level are involved. The variety of roles or effects that stress may have in addition to the motivational one has been discussed by Lazarus, Deese, and Osler (**18**) and more recently by Farber (**7**). Particularly pertinent to the present discussion is the finding that there are wide individual differences in response to such stress, some individuals improving in performance, others decreasing, and still others being unaffected. The direction of the effect of stress has further been related to several personality variables (**18**). The *S*s scoring at the extremes of the MAS continuum may react to such stress with characteristically different patterns as well. Thus, it is possible that with increasing degrees of stress, differences between anxious and nonanxious other than drive may be aroused and become responsible, at least in part, for the discrepancy between the performance levels of such groups.

Unfortunately, the two available studies involving psychological stress do not permit an evaluation of this suggestion (nor of the possibility that stress of any type, physical or psychological, may have a similar effect in tasks of sufficient complexity). Both, it will be recalled, used learning tasks of such a type that an increase in drive level might be expected to result in deterioration of performance. Thus, it could be argued that the anxious were "threatened" (had their drive level increased) by the stress instructions and hence deteriorated in performance in comparison to their neutral control group while the fact that the nonanxious under stress did not show a similar inferiority merely indicates that they were emotionally unaffected by the stress conditions. The only hint that more might be involved than drive level is contained in the Lucas study in which nonanxious improved with a greater number of failure experiences while the anxious became worse. Such a finding suggests further that these additional factors, if any, might act in the direction of interfering with the performance of anxious *S*s and of facilitating the performance of nonanxious. Additional research upon the effects of stress on anxiety groups, particularly with tasks of different levels of complexity is certainly needed to provide information about these possibilities.

The suggestion that at least psychological stress may have other than drive effects on anxious and nonanxious *S*s in complex tasks bears some resemblance to the empirical predictions proposed by Sarason and Mandler and their associates (**22**, **23**, **29**) for the performance of groups selected by a different measuring instrument, a questionnaire of "test anxiety," designed to select individuals react-

ing with different degrees of anxiety to intelligence tests and course examinations. These investigators hypothesized that such high-anxious individuals react to an experimental situation represented as a test of intelligence or the like (thus, according to their conception, creating stress) not only with more anxiety or drive than low-anxious but also, as a result of past learning, have evoked by their anxiety irrelevant response tendencies which interfere with task performance. Under increasing stress (such as reports of failure) the performance of high-anxious *S*s worsens because of the arousal of a greater number of these irrelevant tendencies, offsetting the facilitating effects of drive; the performance of the low-anxious, however, improves with greater stress due to an increasing drive level, unaccompanied by irrelevant tendencies. Such a theory, although predicting the same results as would be expected from the notions being put forward here about the effect of stress on the performance of anxious and nonanxious in complex tasks, differs from these suggestions in several ways. In contrast to drive theory, Sarason and Mandler seem to imply that other things being equal, heightened drive always results in raising performance, independent of the type of task involved. Further they propose that the effect of stress is to evoke certain disruptive response patterns in addition to drive only for high-anxious *S*s while the suggestion of the present writer is that additional factors may be elicited under stress for both anxiety extremes although their effects on performance may be in the opposite direction.

Although Sarason and his colleagues have confined their interests to "test anxiety" and its effects, primarily, on intelligence-test items under stressful conditions, Child (**3**) has proposed that all the work done with *S*s scoring at the extremes on the MAS, independent of whether stress is introduced, could be more plausibly explained by such an interference theory. These task-irrelevant responses are always present in anxious *S*s, as well as a higher drive level, Child states, but they disrupt performance only in complex situations "where the subject is already in conflict between various response tendencies relevant to the task [so that] the presence of irrelevant response tendencies heightens the conflict and interferes with performance to a greater extent than increased drive improves it" (**3**, p. 154).

It would appear to the present writer that a theory that attempts to attribute all inferiority of performance to irrelevant tendencies would either be forced to predict that anxious *S*s would always be inferior to nonanxious in such complex tasks as verbal learning (since it seems hard to maintain that even with verbal materials having little intratask interference, irrelevant extratask responses could not interfere with performance) or, if already obtained results are to be explained, that anxiety level and its correlated irrelevant response tendencies would shift up and down abruptly from task to task and even from stimulus to stimulus within a task as the number of competing response tendencies directly elicited by a stimulus varied. Tieing the number of extratask responses to the number of intratask interferences would seem merely to be adding one more variable to those considered by drive theory without making different predictions in the situations to which drive theory has been thought to be applicable.

It is interesting to note that the suggestions being proposed here concerning the possible role of response as well as drive differences in the performance of anxious and nonanxious *S*s in stress situations leads to a different prediction then do Child's hypotheses in certain cases. According to the present writer, on verbal tasks in which anxious *S*s are demonstrated to be superior to nonanxious under neutral conditions, the introduction of stress might be expected to minimize this difference between groups or even to reverse its direction, the performance of anxious *S*s being lower than under neutral conditions and the nonanxious possibly being higher. Child, while perhaps also expecting nonanxious *S*s to be better under stress than under neutral conditions, would be forced to predict that an anxious group under stress would be the same as or even superior to its neutral control group rather than worse. That is, the fact that under neutral conditions the anxious *S*s perform at a higher level than nonanxious would indicate, according to Child, that this was a situation in which making irrelevant responses does not interfere with task performance, the difference between groups in favor of the anxious being due, then, to their higher drive. While stress might increase the drive level of anxious *S*s and hence the magnitude or number of the task-irrelevant responses, these latter would still not compete with task-relevant responses since the task is the same.

Still another interpretation of the relationship between anxiety and stress has been suggested, the predictions of which are quite opposed to any of those previously discussed. On the basis of their findings with serial learning that the performance of nonanxious groups deteriorated with shock while that for the anxious did not, Deese, Lazarus, and Keenan (**4**) suggested that the MAS measures not so much anxiety as how individuals defend themselves against anxiety, and further, that MAS scores are related to the hysteria-psychasthenia continuum. The latter proposal arose from the finding that (with overlapping items excluded) there was a positive correlation of .40 between the MAS and the Psychasthenia (*Pt*) scale on the MMPI and a −.23 correlation between the MAS and the Hysteria (*Hy*) scale. By assuming that nonanxious *S*s are hysterical individuals who are unable to maintain their defenses in the face of objective inescapable stress (e.g., shock, as opposed to psychological stress), and therefore are greatly disturbed by it while the anxious are psychasthenic and therefore react to objective threat coolly and intellectually, they believe their results become intelligible. The same explanation has been offered by Eriksen (**5**), who found that *S*s scoring high on the *Hy* scale exhibited more stimulus generalization in an investigation involving shock than did high *Pt* *S*s. These results, Eriksen stated, were inexplicable in terms of drive theory. In attempting to evaluate these hypotheses (and leaving aside any questions of the clinical validity of the various measures employed) it might be well to inject a historical note. In developing a scale for the selection of *S*s, the present writer deliberately attempted to include items descriptive of overt or manifest anxiety and avoided including items describing behavior not itself "anxious" but said to be a defense against an internal anxiety precisely because it was the purpose of the scale to select *S*s differing in functioning anxiety

level in the experimental situation; to the extent that defenses were effective in keeping anxiety at a minimum, inclusion of "defense items" on the scale would have been self-defeating.

The conflict between the hypothesis of Eriksen, Deese, *et al.*, and the assumptions made by drive theorists in using the MAS is not whether some individuals scoring low on the scale are potentially anxious individuals with good defenses, but rather whether the introduction of special conditions such as shock so affect a sufficient number of low scoring *S*s as to wipe out or reverse the direction of difference in drive or emotionality between low- and high-scoring groups that exists under neutral conditions. If *S*s are thus affected, drive theorists must either abandon the MAS for a different selective instrument, or restrict themselves to testing groups in situations in which defenses are assumed to be operating.

An examination of the available evidence suggests that no modification of the postulated relationship between anxiety scores and drive level needs be made at the present time (if it is understood that the purpose of drive theory is to investigate the effects of drive once in operation rather than the development of a comprehensive theory of anxiety as a personality phenomenon). That is, the results of Deese *et al.* (**4**) seem deviate; no other investigation involving noxious stimulation (since psychological stress does not assault hysterical defenses) has obtained results that would be expected if the anxiety level of low scoring *S*s increased up to or beyond that of the high scoring *S*s. If such stimulation has any differential effect at all, it appears to be in the direction of increasing the anxiety of the anxious group proportionately more than the nonanxious. Examining the Eriksen results and accepting them as reliable, there seems to be no firm basis for suggesting that drive theory would have predicted more stimulus generalization for the high *Pt* group than the high *Hy*. Such a claim rests on the assumption that all nonanxious *S*s would be low *Hy* and all anxious high *Pt*. The magnitude of the reported correlation coefficients, particularly between the MAS and the hysteria scale does not make this assumption seem too reasonable. Even if high *Hy S*s do become disturbed under nonescapable stress, a sufficient number of *S*s could remain in the nonanxious group who were "genuinely" nonanxious, or whose defenses remained intact, to have a nonanxious group exhibit less stimulus generalization than the anxious. More relevant than such armchair argument, however, are Rosenbaum's (**28**) results. Using an experimental arrangement very similar to Eriksen's, Rosenbaum found, it will be recalled, more stimulus generalization for anxious than nonanxious, and even more important, that the difference between groups was significant *only* under the conditions of strong shock.

MAS and Clinical Measures of Anxiety

As was indicated earlier, the meaning of the term "anxiety" as used in the studies attempting to determine the relationship between drive and performance has been only in terms of MAS scores. While such pure operationism is methodologically sound, the generality of these results would be considerably expanded were a relationship established between the MAS and more common clinical definitions of anxiety. Most valuable

would seem to be a comparison of scale scores with observers' ratings of overt behavior since other diagnostic tests of anxiety are themselves purported to be indicators of such behavior. Fortunately, several studies relating MAS scores and observational data have been carried out. In the first of these investigations, reported by Gleser and Ulett (**9**) of Washington University, a psychiatrist rated 151 normal individuals and 40 psychiatric patients with overt anxiety as a prominent symptom after an hour interview with each subject. Ratings were made on an 8-point scale of anxiety-proneness, defined as the tendency for overt anxiety symptoms to appear in a stressful situation. For the total group the correlation between these ratings and MAS scores was .61. Other similar studies by the Washington group (**45, 46**) with more restricted samples indicated lower coefficients. In a study of 110 male students, involving the judgments of two psychiatrists, the ratings correlated .28 and .29 with MAS scores for the two raters, while the interjudge reliability was .28 (**46**). All correlations were significant. Lastly the Washington group reported a coefficient of .40 between the ratings of a single psychiatrist and anxiety scores for 141 normal *S*s (**45**).

Operating in a student-counseling-center setting, Hoyt and Magoon (**13**) asked experienced counselors to rate their own clients ($N=289$) into one of three groups: high, medium, or low manifest anxiety. Comparing the mean MAS scores for each of the resulting anxiety groups, an extremely significant chi square was found, while the contingency coefficient, used as an estimate of the *r* to be expected if the variable had been continuous, was .47. Using a still different criterion of clinical anxiety, Kendall (**16**) had pairs of nurses rate TB patients on their ward on a 7-point rating scale for each of nine aspects of manifest anxiety. Selecting from the 93 patients so rated the upper and lower 27% in terms of MAS scores, Kendall compared the difference in mean over-all anxiety ratings for the two groups and found it to be statistically insignificant; taking only the upper and lower 13% on the MAS, a very significant *t* between mean ratings was obtained.

Finally, a study by Buss, Wiener, Durkee, and Baer (**2**) represents one of the few investigations utilizing hospitalized psychiatric patients. Each of their 64 patients was interviewed and then rated by four psychologists on nine aspects of directly observed and reported anxiety. Correlations between judges' pooled ratings and MAS scores ranged between .16 to .68 for these various aspects; the correlation with an over-all rating of anxiety was .60.

The variation in the training of the raters, opportunity for observation, rating scales, and populations from which the subjects were drawn makes it difficult to formulate any statement about the "validity" of the MAS. To the extent that all of these observational criteria are themselves correlated and are agreed to be clinically acceptable indices of manifest anxiety, there does seem to be some relationship between MAS and observed behavior. These results suggest, then, that the experimental results obtained with the anxiety scale might also hold for groups selected according to clinical criteria. Such studies as have been reported about the performance of clinically selected anxious groups on comparable tasks tend to confirm this suggestion (**1, 20, 30, 47**).

In addition to the experimental studies of the performance of anxious and nonanxious groups already discussed, a number of other investigations have reported differences in the behavior of anxious and nonanxious *S*s, ranging from indications of number of food aversions (**31**) to performance in problem-solving tasks (**21, 49**). The exclusion of these many experiments from consideration here, due to the limited purpose of this paper—that of assessing the evidence directly relevant to drive theory—points up what has not always been fully appreciated about this theory. It is an extremely restricted one, referring only to the effects of drive level (rather than all characteristics of anxious and nonanxious individuals) in relatively simple learning situations. The major prediction of the theory, that there is an interaction between anxiety level and task complexity, seems to be fairly well substantiated by experimental evidence, although more exact deductions have either not been tested as yet or have not fared as well. Whether the theory can be successfully applied to more complex situations than those for which it originally seemed appropriate, as some have attempted to do, or whether additional variables can be added to it and thus broaden its usefulness remains for future research to determine.

REFERENCES

1. Bitterman, M. E., & Holtzman, W. H. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 615–623.
2. Buss, A. H., Wiener, M., Durkee, A., & Baer, M. The measurement of anxiety in clinical situations. *J. consult. Psychol.*, 1955, **19**, 125–129.
3. Child, I. L. Personality. *Annual Rev. Psychol.*, 1954, **5**, 149–170.
4. Deese, J., Lazarus, R. S., & Keenan, J. Anxiety, anxiety-reduction, and stress in learning. *J. exp. Psychol.*, 1953, **46**, 55–60.
5. Eriksen, C. W. Some personality correlates of stimulus generalization under stress. *J. abnorm. soc. Psychol.*, 1954, **49**, 561–565.
6. Farber, I. E. Anxiety as a drive state. In M. R. Jones, (Ed.), *Nebraska symposium on motivation.* Lincoln, Nebraska; Nebraska Univer. Press, 1954.
7. Farber, I. E. The role of motivation in verbal learning and performance. *Psychol. Bull.*, 1955, **52**, 311–327.
8. Farber, I. E., & Spence, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.
9. Gleser, Goldine, & Ulett, G. The Saslow Screening Test as a measure of anxiety-proneness. *J. clin. Psychol.*, 1952, **8**, 279–283.
10. Gordon, W. M., & Berlyne, D. E. Drive-level and flexibility in paired-associate nonsense-syllable learning. *Quart. J. exp. Psychol.*, 1954, **6**, 181–185.
11. Hilgard, E. R., Jones, L. V., & Kaplan, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–99.
12. Hovland, C. I. Experimental studies in rote-learning theory: III. Distribution of practice with varying speeds of syllable presentation. *J. exp. Psychol.*, 1938, **23**, 172–190.
13. Hoyt, D. P., & Magoon, T. M. A validation study of the Taylor Manifest Anxiety Scale. *J. clin. Psychol.*, 1954, **10**, 357–361.
14. Hughes, J. B., II, Sprague, J. L., & Bendig, A. W. Anxiety level, response alternation, and performance in serial learning. *J. Psychol.*, 1954, **38**, 421–426.
15. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
16. Kendall, E. The validity of Taylor's Manifest Anxiety Scale. *J. consult. Psychol.*, 1954, **18**, 429–432.
17. Lazarus, R. S., Deese, J., & Hamilton, R. Anxiety and stress in learning: the role of intraserial duplication. *J. exp. Psychol.*, 1954, **47**, 111–114.
18. Lazarus, R. S., Deese, J., & Osler, Sonia, F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293–317.
19. Lucas, J. D. The interactive effects of

anxiety, failure, and interserial duplication. *Amer. J. Psychol.*, 1952, **65**, 59–66.
20. Malmo, R. S., & Amsel, A. Anxiety-produced interference in serial rote learning with observations on rote learning after partial frontal lobectomy. *J. exp. Psychol.*, 1948, **38**, 89–101.
21. Maltzman, I., Fox, J., & Morrissett, L., Jr. Some effects of manifest anxiety on mental set. *J. exp. Psychol.*, 1953, **46**, 50–54.
22. Mandler, G., Sarason, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.
23. Mandler, G., & Sarason, S. B. The effect of prior experience and subjective failure on the evocation of test anxiety. *J. Pers.*, 1953, **21**, 336–341.
24. Matarazzo, J. D., Ulett, G. A., & Saslow, G. Human maze performance as a function of increasing levels of anxiety. *J. gen. Psychol.*, 1955, **53**, 79–96.
25. Montague, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–96.
26. Ramond, C. Anxiety and task as determiners of verbal performance. *J. exp. Psychol.*, 1953, **46**, 120–124.
27. Rosenbaum, G. Stimulus generalization as a function of experimentally induced anxiety. *J. exp. Psychol.*, 1953, **45**, 35–43.
28. Rosenbaum, G. Stimulus generalization as a function of clinical and experimentally induced anxiety. Unpublished doctor's dissertation. State Univer. of Iowa, 1950.
29. Sarason, S. B., Mandler, G., & Craighill, P. G. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 561–565.
30. Schiff, E., Dougan, C., & Welch, L. The conditioned PGR and the EEG as indicators of anxiety. *J. abnorm. soc. Psychol.*, 1949, **44**, 549–552.
31. Smith, W., Powell, Elizabeth K., & Ross, S. Manifest anxiety and food aversions. *J. abnorm. soc. Psychol.*, 1955, **50**, 101–104.
32. Spence, K. W. Learning and performance in eyelid conditioning as a function of the intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.
33. Spence, K. W. Current interpretations of learning data and some recent developments in stimulus-response theory. In *Learning theory, personality theory, and clinical research. The Kentucky symposium.* New York: Wiley, 1953.
34. Spence, K. W., & Beecroft, R. S. Differential conditioning and level of anxiety. *J. exp. Psychol.*, 1954, **48**, 399–403.
35. Spence, K. W., & Farber, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.
36. Spence, K. W., & Farber, I. E. The relation of anxiety to differential eyelid conditioning. *J. exp. Psychol.*, 1954, **47**, 127–134.
37. Spence, K. W., Farber, I. E., & Taylor, Elaine. The relation of electric shock and anxiety to level of performance in eyelid conditioning. *J. exp. Psychol.*, 1954, **48**, 404–408.
38. Spence, K. W., & Taylor, Janet A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.
39. Spence, K. W., & Taylor, Janet A. The relation of conditioned response strength to anxiety in normal, neurotic and psychotic subjects. *J. exp. Psychol.*, 1953, **45**, 265–272.
40. Taylor, Janet A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
41. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
42. Taylor, Janet A., & Chapman, J. P. Paired-associate learning as related to anxiety. *Amer. J. Psychol.*, 1955, **68**, 671.
43. Taylor, Janet A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.
44. Taylor, Janet A., & Spence, K. W. Conditioning level in behavior disorders. *J. abnorm. soc. Psychol.*, 1954, **49**, 497–502.
45. Ulett, G. A., Gleser, Goldine, Lawler, A., & Winokur, G. Psychiatric screening of flying personnel. IV. An experimental investigation of development of an EEG index of anxiety tolerance by means of photic stimulation—its validation by psychological and psychiatric criteria. *USAF Sch. Avia. Med. Proj. Rep.*, Aug., 1952, Proj. No. 21-37-002.
46. Ulett, G. A., Gleser, Goldine, Starr, P., Haddock, J., Lingley, L., & Lawler, A. Psychiatric screening of

flying personnel. V. Further studies towards the development of an electro-encephalographic screening technique. *USAF Sch. Aviat. Med. Proj. Rep.*, Aug., 1953, Proj. No. 21-0202-007.

47. Welch, L., & Kubis, J. The effect of anxiety on the conditioning rate and stability of the PGR. *J. Psychol.*, 1947, **23**, 83–91.

48. Wenar, C. Reaction time as a function of manifest anxiety and stimulus intensity. *J. abnorm. soc. Psychol.*, 1954, **49**, 335–340.

49. Wesley, E. L. Perseverative behavior in a concept-formation task as a function of manifest anxiety and rigidity. *J. abnorm. soc. Psychol.*, 1953, **48**, 129–134.

Received August 1, 1955

Psychological Bulletin
Vol. 53, No. 4, 1956

BEHAVIORAL EFFECTS OF IONIZING RADIATIONS

ERNEST FURCHTGOTT[1]
University of Tennessee

Psychology has not been by-passed in the current general interest in ionizing radiations. Since World War II a number of laboratories maintained by the U.S. government have conducted research in this area. In addition, several research projects have been sponsored by government agencies in non-Federal institutions. On March 31, 1955, there were in progress no less than seven such separate projects having no security classification (**50**). All of this activity would seem to warrant a brief review of the problem.

The Stimuli

The biological effects of high energy radiations are ascribable primarily to changes brought about in cells by *ionization*, defined as the removal of electrons from atoms. Different types of radiations produce biological effects differing primarily quantitatively, rather than qualitatively. Two general classes of radiations may be distinguished.

1. *Material radiations* consist of streams of particles which transfer their kinetic energy to the targets which they strike. The particles differing in mass and/or electrical charge are *neutrons*, *alpha particles*, *electrons* (*beta particles*), *deuterons*, or *protons*. These radiations have been utilized only very rarely in behavioral studies.

2. *Electromagnetic radiations* consist of oscillating electric and magnetic fields. They do display also corpuscular (photon) properties. Psychologists are familiar with "light" rays which lie in the frequency range of 10^{14} cycles per second (wavelength range $9\times10^{-5}-4\times10^{-5}$ cm.). Radiations above 10^{16} cycles per second are capable of ejecting inner electrons from atoms. Radiations in the $10^{16}-10^{20}$ cycles per second range ($10^{-6}-10^{-10}$ cm.) are called *X rays*; those between $10^{19}-10^{22}$ cycles per second ($10^{-9}-10^{-12}$ cm.) *gamma rays* (the latter are usually produced by oscillating currents within the atomic nuclei themselves). Gamma rays often accompany the disintegration of radioactive substances.

The relative biological effectiveness of various radiations is a function not only of the total number of ions formed, but also of the spatial distribution of the ions in the tissues. The terms *linear ion density* or *linear energy transfer* are used to express the relative density of ionization per unit length of tissue. Beta and gamma rays produce 6.3–11 ions per micron of tissue, 1,000 kv. X rays approximately 15, 200 kv. X rays 80 and lower voltage X rays a still higher number. Ionization following neutron radiations produces up to 9,000 ions per micron of tissue (**26**, p. 118). Biological effectiveness of radiation increases, decreases, or is independent of linear energy transfer. Thus some activities are affected more by alpha particles than by gamma rays, while in other functions the reverse may be the case. In mammals we usually find that effectiveness increases with ion density.

[1] I wish to express my gratitude to Dr. S. R. Tipton of the U.T. Department of Zoology for critically reading portions of the manuscript.

Measurement of Radiations

Ideally we would like to measure the actual amount of ionization in tissues, but this is not possible. We must be satisfied with specifying the physical characteristics of the source and the target. Ionization of air is actually approximated for ionization in tissues. In the case of X and gamma rays the unit *roentgen, r.*, is defined as that quantity or dose of X or gamma radiation which produces in 0.001293 g. of air one electrostatic unit of ions (**37**, p. 90).

In the case of material radiations a different unit is used, the *roentgen equivalent physical, rep,* which is "that quantity of ionizing radiation which will produce 1.6×10^{12} ion pairs per gram of tissue" (**37**, p. 436). Occasionally, the *roentgen equivalent man, rem,* unit is used which is that "quantity of radiation which when absorbed by man produces an effect equivalent to that produced by absorption of one roentgen of X or gamma radiation" (**37**, p. 436).

A few values will be cited to make the roentgen unit more meaningful. The safe human daily whole-body exposure has been set between 0.05 – 0.25 r. per day (**37**, p. 436; **64**, p. 89). The threshold for the mitotic effect in the grasshopper is 8.0 r. (**64**, p. 89). The 30-day 50 per cent lethal dose after 100–250 kv. X-ray whole-body exposure is about 315 r. for the dog, around 500 r. for man (**55**, p. 930).

General Principles of Radiobiology

It was pointed out previously that radiation-induced effects result primarily from ionization producing physicochemical changes in the living cells. Two general theories concerning the mode of action of radiation have been put forward. According to the *target* theory certain molecules of the cell are especially radiosensitive and it is the change in these specific parts which accounts for the observed radiation effects. Opposing theorists suggest that radiation affects the cell as a whole by releasing certain chemical agents which interfere with the normal cell metabolism. Actually there is evidence supporting both viewpoints. For purposes of this review it is not necessary to examine this problem any further.

The following variables are important in the study of radiation effects: (**55**)

1. *Quantity.* In most instances effects are directly related to the dose.

2. *Rate of delivery or dosage* (sum of doses accumulated over a period of time). In most cases effectiveness of a given dose decreases with a decrease in rate of exposure. Recovery may account for this. For example, in the monkey a single dose of 7,500 r. applied to the spinal cord produces paraplegia, but two daily 5,000 r. doses or five daily 3,000 r. doses are required (**48**).

3. *Type of radiation.* Usually in mammals effectiveness is directly related to the specific ion density of the radiation.

4. *Manner of exposure.* Responses to total-body irradiation are different from those in which only a selected part of the organism is exposed. Shielding of certain parts of the body (spleen, extremities, etc.) can decrease the effectiveness of total-body exposure. This is especially important in the study of the effects on the c.n.s. since doses larger than the median lethal total-body dose are necessary for changes to be apparent.

5. *Time after exposure that observations are made.* Many of the radiobiological effects exhibit latencies. These may be of varying order of magnitudes ranging from seconds to years.

6. *Species differences.* The 30-day LD_{50}[2] for total-body X irradiation for the rabbit is approximately 800 r., for the guinea pig 200–400 r., rat 600–700 r., monkey 500 r. (**55**, p. 930).

7. *Sex differences and individual differences within the same species.* For example, the same dose of X rays kills more male than female mice, but affects the weight of females to a greater extent (**7**).

8. *Conditions of the organism.* Conditions which may be called "stress," i.e., deviation from normal resting state, usually enhance effectiveness of radiations. Vitamin deficiencies, infections, low temperatures in unacclimated animals, exhaustive exercises, adrenalectomies all seem to increase radiation effects.

9. *Drugs and anoxia.* Certain drugs like cysteine, glutathione, alcohol, and anoxia actually depress radiation effects.

10. *Reproductive activity of the tissues.* As early as 1906 Bergonié and Tribondeau hypothesized that proliferating tissues are usually most radiosensitive. We find, for example, that while the nervous system of adult organisms is relatively radioresistant the embryonic neurons are extremely radiosensitive.

Radiation sensitivity varies considerably from tissue to tissue. For a detailed discussion the reader may consult the radiation literature. We shall mention here only a few of the effects, of interest to the psychologist.

The hematopoietic system is extremely radiosensitive. A decrease in the number of circulating lymphocytes is one of the most sensitive indicators of radiation overexposure. Other blood components also show pathological changes. Hemorrhagic manifestations are also quite common after acute irradiation. Vascular changes are major contributors to the brain pathologies observed after large doses of irradiation (**9**). Generalized circulatory changes are only minor after median lethal doses, but with larger doses the effects are more pronounced (**56**).

There is some disturbance in water metabolism. Several studies have reported changes in water intake after X irradiation (**15**, **53**, **54**).

The gastrointestinal tract is extremely radiosensitive. Anorexia, nausea, and vomiting are among the clinical symptoms of radiation sickness (overexposure to radiation). Depression of food intake and a loss of body weight can be observed in irradiated animals. The magnitude and duration of the depression are a function of the dosage (**63**). Loss of body weight can be thus used as an indication of radiation sickness.

The endocrine glands, except for the gonads, are relatively resistant to radiation damage. However, radiations act as "stressors" and they give rise to the well-known pituitary-adrenocortical stress response (**56**).

The cornea, conjunctiva, and the lens of the eye are also quite radiosensitive but the latency of human radiation cataracts may be measured in terms of years (**56**).

Muscle is very resistant to radiation. The nervous system is also relatively radioresistant. Both will be considered in greater detail further on.

Pre- and Neonatal Radiation

An excellent review of the effects of prenatal irradiation has been written by L. B. Russell (**61**).

One of the crucial variables in prenatal irradiation is the stage at which exposure occurs. Russell (**61**) divides

[2] Dose required to kill 50 per cent of the animals during the first 30-day postirradiation period.

the mammalian gestation period into three stages; preimplantation, major organogenesis, period of the fetus. In the rat these periods correspond to the following postconception days: 0–7, 8–15, 16-term. During the preimplantation period radiation produces a high percentage of prenatal deaths, but the survivors are usually normal. The radiation during the period of major organogenesis results in lower prenatal mortality, but it is the most sensitive period for the production of morphological abnormalities. Radiation during the period of the fetus produces lesser changes.

Among the most sensitive systems during the prenatal period is the central nervous system. Russell (**61**) quotes studies dating back to 1907 which show marked morphological changes following X irradiation. In a series of studies on rats and mice Hicks (**29**, **30**) showed that X irradiation during different stages of the gestation period affects different parts of the nervous system. There seem to be critical periods for abnormalities of various types. Irradiation during the first eight days of embryonic life produced no effects on the n.s. of surviving animals. Irradiation on the ninth day resulted primarily in anencephaly; on the tenth day in encephalocele and cerebral deformation; on the eleventh day it narrowed the aqueduct, produced hydrocephaly or encephalocele; on the thirteenth to the sixteenth day the basal ganglia, cortex, hippocampus, and corpus callosum were damaged. From the sixteenth day of gestation through the neonatal period the cerebellum is especially radiosensitive. Hicks emphasizes, however, that the above periods are only indices of the most frequently occurring pathologies and that there is no exact relationship between age of irradiation and specific malformations. Also it should be pointed out that it is rather difficult to determine the precise age of embryos. Wilson and co-workers (**68**, **69**) have shown that neural damage is directly related to dosage. They irradiated rats on the ninth and tenth day of gestation with doses ranging from 25 to 400 r. On the ninth day 25 r. produced ocular malformation in only a small percentage of animals; 50 r. affected 72 per cent of the animals, 100 r. produced anophthalmia, microphthalmia, or other ocular malformations in 90 per cent of the animals; 200 r. proved fatal to all embryos. Brain damages showed similar trends. The data for the animals irradiated on the tenth day of gestation were similar, except that the doses required were higher. Fifty r. had little effect, but 100 r. resulted in anomalous eye development in 75 per cent of the cases.

In this connection a study by Rugh *et al.* (**59**) is of some interest. Rat fetuses 13.5 days old were exposed to 300 r. of X irradiation. In animals examined four hours after exposure the retinae revealed massive damage. On the other hand, animals examined six to seven days after birth had few signs of injury. Apparently a recovery process took place not by repair of the damaged cells, but by proliferation of the more radioresistant precursor neuroectoderm cells.

There are a number of clinical reports of various abnormalities such as microcephaly, hydrocephaly, mental deficiency, ocular malformations, blindness and other types of neural malformations which are ascribed to fetal X irradiation (**25**, **49**). Microcephaly is the most frequently reported abnormality—17 out of 25 abnormal cases in one study (**49**). In some clinical studies, however, no damage is reported following pelvic irradiation during pregnancy (**61**, p. 909). It is possible that the exposure

in the latter cases occurred after the critical period. In the study of 30 pregnant women who showed one or more major signs of radiation following the Nagasaki atomic bomb blast, four out of sixteen surviving children showed signs of mental retardation (**70**). The report does not specify the nature or extent of the deficit.

So far only two studies have been reported which measured specific behavioral consequences of prenatal irradiation. Levinson (**42**) fetally irradiated rats with 300 to 600 r. X rays on the 11th, 13th, 15th, 17th, and 19th postconception days. When the animals were 50 days old they were tested on a Lashley Type III maze. Learning measured in terms of number of trials necessary to reach a criterion, number of errors, and time spent in the maze was impaired with the deficits directly related to the radiation dose. Radiation on the 13th day produced the greatest changes. This agrees roughly with Hicks' timetable for cortical damage (**29, 30**). Variability was larger in the experimental groups than in controls. Tait *et al.* (**65**) X-irradiated rats during the final week of pregnancy using 30, 90, 180, and 360 r. The offspring of the animals receiving 90 or more r. were significantly poorer maze learners than control animals.

Summary. While there is a great deal of evidence for the relative radiosensitivity of the fetal nervous system, our behavioral data are rather scant. We do not know what kinds of activities aside from maze learning are affected nor the lower thresholds of radiation-induced changes. The latter may be of practical significance.

The Adult Nervous System

It has been known for a long time that the adult nervous system is *relatively* radioresistant. Doses in the median total-body lethal range produce no observable neural changes. However, with larger doses, in the case of mammals generally over 1,000 r., a number of investigators have obtained definite signs of neural degeneration in a variety of organisms—man, monkey, dog, rat, rabbit, fish, (**1, 2, 6, 9, 10, 12, 27, 31, 45, 46, 48, 59, 61**). In general the amount of degeneration observed is directly related to the dose and conversely an indirect relationship holds for the latency (**2, 9, 10, 12, 31, 59**). With relatively low doses, a few thousand r., the latency may be a matter of months, a year, or longer (**2, 10, 27, 45**). Many investigators assume that the initially observed neuronal damage is a secondary effect resulting from damage to the vascular system in the brain (**6, 9, 27, 58, 60**). Some recent studies, however, deny the necessity of this assumption (**2**). It might be mentioned here also that because of certain methodological advantages the use of radioactive cobalt has been proposed for the production of circumscribed brain lesions (**62**).

Aside from histological studies, we have information on functional changes. Reflex excitability decreases as a function of dose, with high doses abolishing the reflex completely (**19, 20**). Frequently enhancement precedes the depression (**2, 23, 39**). But again it should be emphasized that median total-body lethal doses produce no easily measurable changes (**13**).

In a study in which the heads of rabbits were irradiated using 12,500 r. (**23**) after a latent period of 30 minutes a convulsive phase with grand mal seizures appeared. This was followed by a somnolent phase of two hours' duration in which the animals were quite inactive. Finally,

in the last stages before death, ataxia was the most pronounced symptom. Changes in equilibrium and disorientation in space have been reported by a large number of investigators (**1**, **10**, **46**, **52**, **58**, **60**). This is in accord with several histological reports that the brain stem and cerebellum are the most frequent sites of radiation necrosis. Hemi- or quadriplegia is a common symptom after large doses (**2**, **10**, **46**, **48**, **58**).

EEG changes have been recorded by several investigators (**2**, **9**, **13**, **39**, **58**), but again the lower threshold is above the median total-body lethal dose. The typical pattern is similar to that seen in seizures, i.e., periodic spikings, high amplitude slow waves.

The most sensitive parts of the brain are the hypothalamus, glial cells, brain stem including the medulla and the cerebellum (**2**, **6**, **9**, **10**, **12**, **31**). The cortex is more radioresistant than these structures, and this is of course significant in behavioral work.

The peripheral nervous system is even less sensitive than the c.n.s. to radiations (**32**). Doses below 10,000 r. are ineffective. It takes 45,000–75,000 r. to abolish nerve conduction in peripheral fibers (**22**). The autonomic n.s. responds with a vagotonia after an initial short duration sympathicotonia (**66**). A slight decrease in pulse amplitude has been reported already after 750 r. Also certain parasympathomimetic effects may be observed during radiation sickness (**56**, p. 996).

Skeletal muscles are also relatively radioresistant. With doses below 6,000 r. no abnormality may be observed (**43**). Gerstner, *et al.* (**21**) applied 50,000 r. to the rabbit gastrocnemius and they noted that fatigue effects could be observed only when high performance was demanded by using a heavy load or a high frequency of stimulation.

Behavioral Changes

Almost since the discovery of X rays investigators have reported various changes in organisms following radiation. Lyman, *et al.* (**46**) in their exhaustive 1933 review of neural changes refer to a study by Tarkhanov who in 1896 observed quieter behavior in flies following X irradiation. There is also an abundance of individual clinical case studies in which radiation was applied for therapeutic purposes. This review will emphasize primarily those studies, however, which were designed specifically to investigate behavioral effects. The latter includes those phenomena customarily included in the field of psychology.

Learning and Performance

The first attempts to assess the effects of radiations on learning were performed in Pavlov's laboratory. Nemenow (**51**, **52**) irradiated the head of one dog with a dose of 1,500 r. There was only a slight drop in his salivary CR's. After an additional 2,200 r., however, the CR's practically disappeared for a period of five weeks. A second dog received 3,500 r. then again 2,800 r. and the results were essentially similar to those seen in the first animal. Lyman, *et al.* (**46**) X-irradiated the occipital part of the head of four dogs with massive doses of 17,000–18,000 r. after their CR's had been stabilized. All of the animals showed a temporary decrease in their salivary CR's, but the onset and duration of this decrement varied. Two of the animals ("excited types") actually showed a rise in CR's preceding the drop. The strength of the responses also varied as a function of the type of CS. One of the animals kept alive for six months after the treatment exhibited a second lowering of CR's following the recovery from the first

decrease. This latency in the gross pathological manifestations is consistent with the other investigations discussed in the previous section. The change in the CR's occurred during a period when the *S* exhibited ataxia, impaired vision, circus movements, and general deterioration of behavior. It was difficult to test the dog. Interpretation of the data from the whole study is obscured by the observation that in three *S*s not only the CR's but the UR's also showed a drop.

In a study for which an abstract only is available, Harlow (**28**) reports that radon tubes inserted into the cortex of ten rhesus monkeys produced progressive loss on delayed reaction, patterned string tests, and simple position habits. No data are given for the dosage used. It was probably quite large in view of other negative findings reviewed below.

No further work was done in this field until after World War II. Furchtgott (**16**) tested rats exposed to 200–500 r. of total X radiation in a four-unit water maze. Neither acquisition nor retention using several criterion measures was affected by the treatment. Arnold (**3**) exposed the heads only of rats to 300–800 r. and tested them for retention of a 14-unit T-maze habit and other irradiated *S*s were tested for the learning of the habit. No statistically significant changes were found. Fields (**14**) studied performance on elevated runways, 32- and 40-choice-point elevated T-mazes, and a 10-choice-5-stage vertical maze of some 500 male rats which had received doses ranging from 100–1,000 r. On the whole radiation had little effect on the performance of the animals except for a decrease in the speed and amount of activity immediately following irradiation which was probably due to the general radiation malaise. Davis (**11**) tested rhesus monkeys in the Wisconsin General Test apparatus following X irradiation but was unable to find any impairment of performance on discrimination-type tasks. In a series of studies sponsored by the U.S. Air Force School of Aviation Medicine (**34**, **57**), monkeys were tested on acquisition, retention, and transfer of multiple discrimination problems immediately and 150 days after exposure to sublethal and lethal doses of X rays. Again the reported results failed to demonstrate any deleterious effects. The only deficit that was noted was an increase in reaction time.

Garcia *et al.* (**18**) established a conditioned aversion to a saccharine solution which was associated with exposure to gamma irradiation. Experimental animals had saccharine solutions in their cages while being exposed for six hours in the gamma field, while control *S*s had tap water. Preference was then tested for 63 postirradiation days. The control group showed no loss of their natural preference for saccharine, while experimental *S*s exposed to only 30 r. showed a significant drop in their saccharine intake. The authors hypothesize a general behavior disturbance during radiation which became associated with the taste stimmuli. It should be pointed out that the animals were being exposed at a very slow rate and some of the general radiation malaise might have been effective for a sufficient length of time for the conditioning. The effectiveness of the low dose used is surprising, however.

Jones *et al.* (**33**) measured the effects of 200–1,000 r. of whole-body X irradiation on activity-wheel performance, using 194 rats. Data were analyzed separately for animals who survived the eight-week experimental period and those that succumbed to radiation injury. Rats which died during the first nine postirradiation

days showed a gradual decrease in activity until death. Those that survived nine days, but died subsequently showed a decrease immediately after irradiation followed by a recovery and a second depression of activity prior to death. All of the surviving animals (doses 200–680 r.—all animals with higher doses died) showed a decrease in activity postirradiation. The 200–300 r. groups recovered completely by the fifth day. The higher-dose groups also showed a partial recovery during the first postirradiation week, but they exhibited a second depression during the third week. The 400–450 r. groups attained normal levels of activity four weeks after irradiation, the 681 r. groups after eight weeks. In general there was a direct relationship between degree and duration of activity depression.

In another study the same group of investigators (**36**) tested the effects of 300–1,000 r. X irradiation on exhaustive swimming exercise. The rats were placed into a 24-gallon tank where they were forced to swim until they were exhausted and sank, remaining below the surface of the water for longer than 30 seconds at which time they were retrieved. Length of swimming time before sinking was measured. Following radiation, performance gradually decreased, reaching a minimum level during the third to fourth postirradiation week. From then on there was a gradual return to the normal level which was attained by the ninth week. While the depression was directly related to dose, the 300 r. group barely differed from control animals. The 500 r. group, however, showed a significant drop and the higher r. animals in turn differed significantly from the 500 r. group.

Furchtgott (**15**) subjected adolescent rats to 300 and 500 r. of X rays and tested their swimming speed in a 12 ft. straight-away tank for 13 days. The 300 r. group did not differ from the controls, but the survivors in the 500 r. group were significantly slower.

Vogel (**67**) daily irradiated with 50 r. X rays six aggressive mice each of whom, prior to the treatment, always defeated submissive animals. Even after irradiation the aggressive animals continued to be dominant until shortly before death.

McDowell (**47**) observed a reduction in "other-animal involved" behavior and visual attention to the activity of other animals following 400 r. of X irradiation in 10 rhesus monkeys. The animals also showed fewer instances of aggression and a greater incidence of lethargy. All of these symptoms are easily understandable considering the general malaise which is associated with radiation.

Leary and Ruch (**38**) exposed 18 rhesus monkeys to 200–400 r. of total-body X irradiation. Cage-crossings were not affected. On the first postirradiation day only for the 400 r. animals (the others were not observed) scratching, grooming, and other signs of activity were depressed—a sign of general malaise. Mechanical puzzle manipulation did not produce statistically significant differences between pre- and postirradiation periods. Pedometer manipulation was impaired in the 400 r. animals (others were not tested) and surprisingly weight-pulling, supposedly a measure of general strength, did not decrease in all animals.

In general it may be said that radiation produces a certain amount of depression in activity which should be most apparent when motivation is low or when the task requires a great deal of effort as in the exhaustive swimming experiment (**36**). The latter effect would tend to parallel Gerstner's, *et al.* (**22**), findings on the

effect of X irradiation on muscular contraction.

There is a puzzling report of a clinical study of 120 patients who had received several single doses of 30–50 r. during a 7–10 day period totaling 150–250 r. of diencephalic X irradiation (**4**). Immediately following the treatments the typical symptoms were numbness, apathy, and tingling sensations in the head region. The day after the irradiation, however, most patients reported spontaneously that they felt euphoric, active, and generally tranquil. This state lasted from a week to several months. Most of the treated patients were neuropsychiatric cases with diagnoses of urticaria, migraine, depression, etc. However, in addition, two medical collaborators subjected themselves to 100 r. administered to the diencephalon and they also experienced the same changes as the patients. Sixty-one of the patients reported changes in their sleep patterns. The sleep on the night following the treatment was usually characterized as "extremely deep," "heavy," or "leaden." In addition 37.5 per cent of the *S*s reported sexual changes, notably improvement in libido, potency, and the menses.

The authors ascribe these changes to hypothalamic stimulation primarily of the anterior, parasympathetic nuclei, a finding in accord with the frequently reported radiation-induced vagotonia (**66**). These results, if confirmed by other investigators, should have therapeutic implications. They also raise many questions of interest to the experimentalist working with animals since we have practically no data on emotional behavior following radiation.

Summary. The lack of any dramatic changes in learning functions following sublethal or just lethal total-body X irradiations reported by several experimenters agrees very well with similar neurophysiological observations on the resistance of the nervous system in that dose range. It takes doses which are well above the median total-body lethal range to produce any neural changes and then there is usually a considerable latency. In the one study in which there was an 18-month time lapse between the treatment and testing, no drastic decrements took place (**14**). Whether a longer period would have any effects is an open question. While acquisition, retention, or transfer are not affected, performance indices which utilize gross muscular activity are impaired to some extent and this impairment persists for a number of months (in the study of swimming endurance [**36**] up to nine months for the most heavily irradiated group). Another factor to be considered is what might be called, for the lack of a better name, general malaise, which includes a lack of motivation to respond to stimuli or initiate activity which is present immediately following radiation and appears again during the second week in more heavily irradiated animals. This is accompanied also by a loss in appetite and drop in body weight.

Sensory Functions

Hearing. In the clinical literature there are reports of improved hearing following X irradiation. In the early thirties Girden (**24**) working in Culler's laboratory attempted to investigate this problem using dogs in the classical conditioning setup. Standard psychophysical procedures were employed to obtain absolute intensity thresholds. Subsequently the heads of twelve animals were irradiated. The study was exploratory in nature and there was no systematic design to test radiation factors. Eight animals were irradiated using 80–100 kv. peak voltage and 5 ma., while four animals got roentgen rays gen-

erated at 200 kv. peak and 5 ma. One animal received 5 r. every day for five months, one 5 r. for four days, one anywhere from 100–1,100 r. on seven days, spaced one to seven days apart, and so forth. The total dosage varied from 20–11,100 r. The animals which were irradiated with the 80–100 kv. rays all showed a transient gain in acuity which averaged 5.5 decibels after a latent period of seven to eleven days. Dosage was apparently not involved since the changes appeared even after the surprisingly low value of 20 r. None of the *S*s irradiated at 200 kv. showed any improvement in acuity.

In a second study Brogden and Culler (**5**) examined more critically the effect of dose and also the frequency variable. Ten animals were irradiated at nine different intensities ranging from 75 to 675 r. The gain in acuity was independent of the dose, varying from 3.84 db to 7.87 db. The duration of the gain was also independent of the dose and it varied from 8.0 to 10.3 days. The latency, however, was inversely related to the dose. At 75 r. it was 7.6 days and at 675 r. it was 2.6 days.

To explore the mechanism of the phenomenon, two dogs were hypophysectomized and irradiated; auditory tests were conducted on a diabetic subject and a normal dog when blood-sugar levels were high and low (by injection of insulin); and blood-sugar levels were measured before and after irradiation in one dog. In all of these cases hypoglycemia was associated with lower auditory thresholds. The authors hypothesize that low sugar levels lessen density and viscosity of cochlear fluids, and thereby decrease resistance to incoming vibrations, and perhaps also the ionic conditions in the cochlea affect the magnitude of the cochlear potentials.

Vision. Fields (**14**) found no effects on brightness or acuity discrimination in rats following X irradiation. Russian workers (**35**) have reported that dermal X irradiation increases the threshold to dark adaptation and that this effect persists for several days. Lenoir (**41**) tested dark adaptation in 11 patients following therapeutic irradiation. In all cases there was a decrease in dark adaptation which was independent of the dose (2,400–6,240 r.). The changes could be detected for 20 to 36 days. The author ascribes this reduction in dark vision to a drop in vitamin A concentration which follows the X irradiation. Furchtgott (**17**) tested brightness discrimination in a Lashley jumping box under conditions of low illumination following 369–469 r. of X irradiation. The performance of the irradiated rats was slightly inferior to that of control animals. It should be noted here also that Cibis *et al.* (**8**) found that rod cells are considerably more radiosensitive than cones. Destruction of rods required 1,700–2,000 r. while the threshold for cones is 10,000–30,000 r.

The work on cataract formation has been reviewed adequately (**40**) and it is omitted here since the studies involve primarily morphological changes.

Other senses. The work on other senses is scant. Lindemann (**44**) observed fifteen patients who received therapeutic X-ray treatments for tumors in the oral cavity. Taste sensitivity and in some cases odor sensitivity were depressed for several months. In an unpublished study Furchtgott found some indication of lowered thresholds to electric shock in rats following sublethal doses of whole-body X irradiation.

Summary. While there is some evidence for changes in sensory functions notably hearing and scotopic vision after irradiation, the available data are quite limited. Much more

research is necessary in the various sensory modalities to determine the factors, if any, which affect perception.

Summary and Conclusions

The published studies pertaining to the behavioral effects of high-energy radiations were reviewed. More studies have actually been performed in this area. The author knows of several additional ones, performed by himself and by others, but the negative results have discouraged the workers from publishing them.

Underlying any discussion of the behavioral effects of radiation is the relative radioresistance of the adult nervous system. Total-body doses in the median lethal range do not seem to produce any gross neural dysfunctions. Except for the instances in which the body is shielded and the radiations are applied to the head only, death will intervene long before any neural changes can be observed. Thus we will not find any significant behavioral changes in those activities which are mediated directly by the nervous system. We have reviewed several studies of learning by different investigators which seem to bear this out. Actually it is possible that an investigation will show a decrement in learning following radiation. However this would be primarily a reflection of the change in the nonassociative learning factors, i.e., motivation and perception of the stimuli.

We have pointed out that radiation produces changes in the blood and body fluids, gastrointestinal tract, and some of the endocrine secretions. Thus the homeostatic energy-controlling mechanisms are affected and we should find, therefore, changes in motivation and performance. We have indeed seen that some of these functions are altered. Food and water intake, exhaustive swimming exercise, activity wheel and pedometer performance, and social behavior changes have been reported. There are still a number of problem areas here such as emotionality, motivation, other than hunger and thirst, which have not been investigated. Here we should mention again that radiation seems to lead to the pituitary-adrenocortical stress reaction and that the hypothalamus and the autonomic n.s. are relatively more sensitive than the cortex. It would seem also that performance which requires a large expenditure of energy or where extrinsic incentives are very small will be affected the most by radiations.

In the sensory field some experimental work has been reported on hearing and vision and we have also clinical data on these and other modalities. On the whole, however, there are large gaps here.

The great sensitivity of the developing nervous system was briefly discussed. The quantity of behavioral data does not approach our knowledge of morphological changes. We have only two studies on maze learning in rats. It would seem that this area should be explored in greater detail and functions other than maze learning could be explored.

The genetic aspects of radiation were not considered since we have no data here on variables which are conventionally classified as psychological.

References

1. Andrews, H. L. & Kirkland, C. B. Terminal phenomena associated with massive doses of X-rays. *Amer. J. Physiol.*, 1953, **175**, 138–140.

2. Arnold, A., Bailey, P. & Laughlin, J. S. Effect of betatron radiations on the brain of the primates. *Neurology*, 1954, **4**, 165–178.

3. ARNOLD, W. J. Maze learning and retention after X radiation of the head. *J. comp. physiol. Psychol.*, 1952, **45**, 358–361.

4. BIRKNER, R., & TRAUTMANN, J. Über die Abhängigkeit psychischer, Schlaf- und genitaler Funktionen von den vegetativen Steuerungszentren im Hypothalamus und die Beeinflussbarkeit dieser Funktionen durch Röntgenbestrahlungen des Zwischenhirngebietes mit kleinen Dosen. *Strahlentherapie*, 1953, **91**, 321–350.

5. BROGDEN, W. J., & CULLER, E. Increased acoustic sensitivity in dogs following roentgen-radiation of the hypophysis. *Amer. J. Physiol.*, 1937, **119**, 13–23.

6. CAMPBELL, B., & NOVICK, R. Effects of beta rays on central nervous tissues. *Proc. Soc. exp. Biol., N.Y.*, 1949, **72**, 34–38.

7. CHAPMAN, W. H. The weight and mortality response of male and female mice in the lethal X-ray dose range. *Radiation Res.*, 1955, **2**, 502–511.

8. CIBIS, P. A., NOEL, W. K., & EICHEL, B. Ocular effects produced by high-intensity X-radiation. *A.M.A. Arch. Opthal.*, 1955, **53**, 651–663.

9. CLEMENTE, C. D., & HOLST, E. A. Pathological changes in neurons, neuroglia, and blood-brain barrier induced by X-irradiation of heads of monkeys. *A.M.A. Arch. Neurol. Psychiat.*, 1954, **71**, 66–79.

10. DAVIDOFF, L. M., DYKE, C. G., ELSBERG, C. A., & TARLOV, I. M. The effect of radiation applied directly to the brain and the spinal cord. I. Experimental investigations on Macacus rhesus monkeys. *Radiology*, 1938, **31**, 451–463.

11. DAVIS, R. T. The performance of rhesus monkeys on food rewarded discrimination type tasks before and after X-radiation. Paper read at Amer. Psychol. Ass., New York, September, 1954.

12. DAVISON, C. & ELLINGER, F. Radiation effects on nervous system and roentgen-pigmentation of goldfish. *Proc. Soc. exp. Biol., N. Y.*, 1942, **49**, 491–495.

13. ELDRED, E., & TROWBRIDGE, W. V. Neurological and EEG findings in the monkey after total body X-irradiation. *EEG clin. Neurophysiol.*, 1953, **5**, 259–270.

14. FIELDS, P. E. The effect of whole body X-radiation upon the psychological abilities of white rats. *Amer. Psychologist*, 1955, **10**, 415–416. (Abstract)

15. FURCHTGOTT, E. Behavioral effects of total X-irradiation in the white rat: an exploratory study. Unpublished doctor's dissertation, Univer. of California at Los Angeles, 1950.

16. FURCHTGOTT, E. Effects of total body X-irradiation on learning: an exploratory study. *J. comp. physiol. Psychol.*, 1951, **44**, 197–203.

17. FURCHTGOTT, E. The effects of X-irradiation on brightness discrimination. *J. Psychol.*, 1952, **34**, 37–41.

18. GARCIA, J., KIMELDORF, D. J., & KOELLING, R. A. Conditioned aversion to saccharin resulting from exposure to gamma radiation. *Science*, 1955, **122**, 157–158.

19. GEĬNISMAN, ÎA. I., & ZHIRMUNSKAÎA, E. A. [On the mechanism of action of Roentgen rays on reflex excitability.] *Fiziol. Zh. S.S.S.R.*, 1952, **38**, 312–318.

20. GERSTNER, H. B., KONECCI, E. B., & TAYLOR, W. F. Effect of local brain X-irradiation on the pinna reflex of guinea pigs. *Radiation Res.*, 1954, **1**, 262–269.

21. GERSTNER, H. B., LEWIS, R. B. & RICHEY, E. O. Early effects of high intensity X-radiation on skeletal muscles. *J. gen. Physiol.*, 1954, **37**, 445–459.

22. GERSTNER, H. B., ORTH, J. S., & RICHEY, E. O. Effect of high-intensity X-radiation on velocity of nerve conduction. *Amer. J. Physiol.*, 1955, **180**, 232–236.

23. GERSTNER, H. B., PICKERING, J. E., & DUGI, A. J. Sequelae after application of high-intensity X-radiation to the head of rabbits. *Radiation Res.*, 1955, **2**, 219–226.

24. GIRDEN, E. Effect of roentgen rays upon hearing in dogs. *J. comp. Psychol.*, 1935, **20**, 263–290.

25. GOLDSTEIN, L. Radiogenic microcephaly; a survey of nineteen recorded cases with special reference to ophthalmic defects. *Arch. Neurol. Psychiat.*, 1930, **24**, 102–115.

26. GRAY, L. H. Comparative studies of the biological effects of X-rays, neutrons, and other ionizing radiations. In N. Howard-Jones (Ed.) *Applied biophysics: Survey of physical methods used in medicine.* Brooklyn, N.Y.: Chemical Publishing Co., 1949. Pp. 114–137.

27. GREENFIELD, M. M., & STARK, M. F. Post-irradiation neuropathy. *Amer. J. Roentgenol.*, 1948, **60**, 617–622.

28. HARLOW, H. F. Effect of cortical lesions induced by radium emanations upon the adaptive behavior of the rhesus monkey. *Psychol. Bull.*, 1933, **30**, 737–738.

29. HICKS, S. P. Acute necrosis and malformation of developing mammalian brain

caused by X-ray. *Proc. Soc. exp. Biol., N.Y.*, 1950, **75**, 485–489.

30. Hicks, S. P. Mechanism of radiation anencephaly, anophthalmia, and pituitary anomalies, repair in the mammalian embryo. *Arch. Path.*, 1954, **57**, 363–378.

31. Hicks, S. P., & Montgomery, P. O'B. Effects of acute radiation on the adult mammalian central nervous system. *Proc. Soc. exp. Biol., N. Y.*, 1952, **80**, 15–18.

32. Janzen, A. H., & Warren, S. Effect of roentgen rays on peripheral nerve of rat. *Radiology*, 1942, **38**, 333–337.

33. Jones, D. C., Kimeldorf, D. J., Rubadeau, D. O., Osborn, G. K., & Castanera, T. J. Effect of X-irradiation on performance of volitional activity by the adult male rat. *Amer. J. Physiol.*, 1954, **177**, 243–250.

34. Kaplan, S. J., & Gentry, G. The effect of sublethal dose of X-irradiation upon transfer of training in monkeys. *USAF Sch. Aviat. Med. Proj. Rep.*, 1953, Proj. no. 21–3501–0003 (Rep. No. 4).

35. Kekcheew, K. KH. [On the changes of thresholds in man in achromatic vision subjected to the action of ultra-short waves, ultra-violet, and X-rays.] *Probl. fiziol. Optiki*, 1941, **1**, 77–79.

36. Kimeldorf, D. J., Jones, D. C., & Castanera, T. J. Effect of X-irradiation upon the performance of daily exhaustive exercise by the rat. *Amer. J. Physiol.*, 1953, **174**, 331–335.

37. Lapp, R. E., & Andrews, H. L. *Nuclear radiation physics.* New York: Prentice-Hall, 1948.

38. Leary, R. W., & Ruch, T. C. Activity, manipulation drive, and strength in monkeys subjected to low-level irradiation. *J. comp. physiol. Psychol.*, 1955, **48**, 336–342.

39. Lee, J. C., & Snider, R. S. Acute radiation effects on electrical activity of cerebrum and cerebellum. *Radiation Res.*, 1954, **1**, 503–504. (Abstract)

40. Leinfelder, P. J., Evans, T. C., & Riley, E. Production of cataracts in animals by X-rays and fast neutrons. *Radiology*, 1955, **65**, 433–438.

41. Lenoir, A. Adaptation und Röntgenbestrahlung. *Radiol. clin., Basel*, 1944, **13**, 264–276.

42. Levinson, Billey. Effects of fetal irradiation on learning. *J. comp. physiol. Psychol.*, 1952, **45**, 140–145.

43. Lewis, R. B. Changes in striated muscle following single intense doses of X-rays. *Lab. Invest.*, 1954, **3**, 48–55.

44. Lindemann, B. Zur Frage der Radiosensibilität des peripheren Nervensystems. *Fortsch. Röntgenstr.* 1949, **71**, 988–993.

45. Lowenberg-Scharenberg, K., & Bassett, R. C. Amyloid degeneration of the human brain following X-ray therapy. *J. Neuropath.*, 1950, **9**, 93–102.

46. Lyman, R. S., Kupalov, P. S., & Scholz, W. Effect of roentgen rays on the central nervous system: Results of large doses on the brains of adult dogs. *Arch. Neurol. Psychiat.*, 1933, **29**, 56–87.

47. McDowell, A. A. The immediate effects of single dose whole body X radiation upon the social behavior and self-care of caged rhesus monkeys. *Amer. Psychologist*, 1954, **9**, 423. (Abstract)

48. McLaurin, R. L., Bailey, O. T., Harsh, G. R., & Ingraham, F. D. The effects of gamma and roentgen radiation on the intact spinal cord of the monkey; an experimental study. *Amer. J. Roentgenol.*, 1955, **73**, 827–835.

49. Murphy, D. P. Outcome of 625 pregnancies in women subjected to pelvic radium or roentgen rays. *Amer. J. Obstet. Gynec.*, 1929, **18**, 179–189.

50. National Science Foundation, Memorandum sent to chairmen of departments of psychology, May 31, 1955.

51. Nemenow, M. I. The effect of roentgen-ray exposures of the cerebral cortex on the activity of the cerebral hemispheres. *Radiology*, 1934, **23**, 86–93.

52. Nemenow, M. I. The effect of roentgen rays on the brain: Experimental investigation by means of the conditioned reflex method. *Radiology*, 1934, **23**, 94–96.

53. Nims, L. F., & Sutton, E. Weight changes and water consumption of rats exposed to whole-body X-irradiation. *Amer J. Physiol.*, 1952, **171**, 17–21.

54. Nims, L. F., & Sutton, E. Adrenal cholesterol, liver glycogen and water consumption of fasting and X-irradiated rats. *Amer. J. Physiol.*, 1954, **177**, 51–54.

55. Patt, H. M., & Brues, A. M. The pathological physiology of radiation injury in the mammal. I. Physical and biological factors in radiation action. In A. Hollaender (Ed.) *Radiation biology.* Vol. I: *High energy radiation.* New York: McGraw-Hill, 1954. Pp. 919–958.

56. Patt, H. M., & Brues, A. M. The pathological physiology of radiation injury

in the mammal. II. Specific aspects of the physiology of radiation injury. In A. Hollaender (Ed.) *Radiation biology.* Vol. I: *High energy radiation.* New York: McGraw-Hill, 1954. Pp. 959–1028.

57. Rogers, C. M., Kaplan, S. J., Gentry, G., & Auxier, J. A. Some effects of cumulative doses of X-radiation upon learning and retention in the rhesus monkey. *USAF Sch. Aviat. Med. Proj. Rep.*, Proj. No. 21-3501-0003 (Rep. No. 11).

58. Ross, J. A. T., Leavitt, S. R., Holst, E. A., & Clemente, D. C. Neurological and electroencephalographic effects of X-irradiation of the head in monkeys. *A.M.A. Arch. Neurol. Psychiat.*, 1954, **71**, 238–249.

59. Rugh, R., & Wolff, J. Resilience of the fetal eye following radiation insult. *Proc. Soc. exp. Biol., N.Y.*, 1955, **89**, 248–253.

60. Russell, Dorothy S., Wilson, C. W., & Tansley, K. Experimental radionecrosis of the brain in rabbits. *Jour. Neurol. Neurosurg. Psychiat.*, 1949, **12**, 187–195.

61. Russell, Liane B. The effects of radiation on mammalian prenatal development. In A. Hollaender (Ed.) *Radiation biology.* Vol. I. *High energy radiation.* New York: McGraw-Hill, 1954. Pp. 861–918.

62. Settlage, P. H., & Bogumill, G. P. Use of radioactive cobalt for the production of brain lesions in animals. *J. comp. physiol. Psychol.*, 1955, **48**, 208–210.

63. Smith, D. E., & Tyree, E. B. Influence of X-irradiation upon body weight and food consumption of the rat. *Amer. J. Physiol.*, 1954, **177**, 251–260.

64. Spear, F. G. The biological effects of penetrating radiations. In N. Howard-Jones (Ed.) *Applied biophysics: Survey of physical methods used in medicine.* Brooklyn, N. Y.: Chemical Publishing Co., 1949. Pp. 83–113.

65. Tait, C. D., Jr., Wall, P. D., Balmuth, M., & Kaplan, S. J. Behavioral changes following radiation. II. Maternal behavior and maze performance. *USAF Sch. Aviat. Med.* (Spec. Rep. No. 11).

66. Vieten, H. Der strahlenbiologische Reaktionsablauf im vegetativen Nervensystem. *Strahlentherapie*, 1949, **79**, 13–58.

67. Vogel, H. H., Jr. The effect of X irradiation on fighting behavior in male mice. *Anat. Rec.*, 1950, **547**. (Abstract)

68. Wilson, J. G., Jordan, H. C., & Brent, R. L. Effects of irradiation on embryonic development. II. X-rays on the ninth day of gestation in the rat. *Amer. J. Anat.*, 1953, **92**, 153–187.

69. Wilson, J. G., & Karr, J. W. Effects of irradiation on embryonic development: I. X-rays on the 10th day of gestation in the rat. *Amer. J. Anat.*, 1951, **88**, 1–33.

70. Yamazaki, J. N., Wright, S. W., & Wright, P. M. A study of the outcome of pregnancy in women exposed to the atomic bomb blast in Nagasaki. *J. cell. comp. Physiol.*, 1954, **43**, Suppl. 1, 319–328.

Received October 31, 1955.

Psychological Bulletin
Vol. 53, No. 4, 1956

COMMENTS ON MEEHL AND ROSEN'S PAPER[1]

SAMUEL KARSON and SAUL B. SELLS
USAF School of Aviation Medicine, Randolph Field, Texas

The recent paper by Meehl and Rosen (**3**) presents a rationale for evaluating the predictive efficiency of psychometric instruments which should be of interest and importance in clinical and personnel research. The purpose of this comment is to emphasize the principle of the dependence of statistical criteria on administrative policy in selecting the appropriate criterion among the cases which they have effectively presented.

The basic reference statistic for evaluation for predictive efficiency is the base rate, according to Meehl and Rosen (**3**, p. **194**). Evaluation of any predictor requires comparison of results based on prediction with the base rates prevailing in the situation. Thus, if one thousand candidates were available for military service and the base rate of noneffectiveness were 5%, 950 successful candidates might be expected without screening. Now, if a screening device operated to admit less than 950 successful candidates in the same situation, Meehl and Rosen would consider such a test less efficient than the base rate.

Their analysis considers three separate cases. The first is *efficiency in detecting cases of poor adjustment.* Here they classify as errors of prediction only the false-positives rejected. When the false-positive rate is higher than the base rate of noneffectiveness, they would conclude that use of the screening test would be less efficient than no screening at all. The second case is *efficiency in prediction for all cases.* Here they classify as errors of prediction both the false-positives rejected and the false-negatives accepted. When the number of successful cases attained through a sample of available individuals is lower as a result of screening than could be expected according to the prevailing base rate, they would consider such screening inefficient. The third case is called *efficiency in detecting cases of good adjustment.* Here only false-negatives are regarded as errors. Thus to the extent that the proportion of successfuls in the sample accepted is greater than expected according to the base rate, they would consider screening to be efficient. They point out, however, that such efficiency is relative, inasmuch as it purchases increased efficiency of personnel accepted at the cost of rejecting some potentially successful candidates in the screening process.

Although the point is implied by Meehl and Rosen, it seems important to emphasize as a general principle that the choice of the appropriate test of efficiency depends on the policies in effect and the purposes of screening required to fulfil them. Widespread misunderstanding of this principle could seriously impair the status of many useful screening and prediction programs. All too often scientists are too preoccupied with considerations of validity, while they fail to recognize the practical problems facing administrators who utilize psychometric techniques. On the other hand, administrators need to understand this principle so that they may avoid the error of rejecting use-

[1] The writers wish to express their appreciation to Dr. Samuel Fulkerson for contributing to the discussion which culminated in the present paper.

ful methods as well as that of accepting inefficient ones, through faulty evaluation.

With specific reference to induction screening of military personnel, the manpower administrator is concerned with supply and demand issues on one hand, and with the burden of additional administration and loss of productive work due to noneffectiveness on the other. In times of manpower scarcity, he may be pressed to utilize every available man. Under such circumstances, he would seek to admit the maximum number from the pool available. Then the Meehl-Rosen Case 2 would be properly applied in evaluating prospective screening devices.

If, however, manpower shortages were less pressing, or if the waste attributable to noneffectiveness were considered too great, the administrator might be agreeable to the rejection of some potentially successful individuals by a screening device which could assure a greater proportion of successful candidates from the number admitted than might be expected according to the base rate. The gross number of successful candidates for any available sample would be less, depending upon the rejection rate for the particular screening device, but the noneffectiveness rate might be reduced. In these circumstances Case 3 would be appropriate to evaluate the increase in proportion of successful candidates as a result of screening and Case 1 could be used to evaluate the cost in terms of false-positive rate.

The criterion implied in Case 2 requires maximization of the number of successfuls in relation to the total pool available, whereas Case 3 requires maximization of the number of successfuls in relation to the number admitted. Both need to be evaluated against the base rate. The former criterion may be dictated in circumstances of manpower scarcity, while the other would reflect a policy decision more sensitive to the cost of accepting and caring for noneffective individuals in hospitals, guardhouses, and nonproductive jobs.[2] Policy, not mathematical reasoning, must dictate the appropriate criterion of evaluation and the proportion of incorrect predictions which can be accepted.

The writers feel that in view of the general excellence of Meehl and Rosen's paper, their oversight in connection with their discussion of Case 1 should be mentioned. They demonstrate (**3**, p. 195) that the use of the Danielson and Clark (**2**) screening inventory would result in a decrease in the total percentage of correct predictions made (from 95% to 79.7%) when comparing the test with the base rates. They do not, however, indicate that the screening inventory has actually succeeded in raising the percentage of correctly predicted "fails" from 5% (base rate) to 13%. Later they do recognize this kind of gain when they demonstrate (**3**, p. 204) that a certain cutting score on the Glueck prediction index succeeds in correctly identifying delinquents with an accuracy of 92.6% as compared with an expected 20% base rate, even though predictions are made for only 2.4% of the population.

[2] It is of interest to note that the current induction screening policy of the armed services emphasizes the second criterion described (**1**, **4**, **5**).

REFERENCES

1. Army Regulation 40–115, *Medical Department physical standards and physical profiling for enlistment and induction*, Dept. Army, 20 August 1948.

2. Danielson, J. R., & Clark, J. H. A personality inventory for induction screening. *J. clin. Psychol.*, 1954, **10**, 137–143.
3. Meehl, P. E., & Rosen, A. Antecedent probability and the efficiency of psychometric signs, patterns or cutting scores. *Psychol. Bull.*, 1955, **52**, 194–216.
4. *Psychiatric Unit Operational Procedures.* Bur. Med. Surg. (ONR Contract No. N6-ori-158.)
5. *Psychiatric Unit Operational Procedures* (Rev. ed., 1954). Neuropsychiat. Br.; Profess. Div., Bur. Med. Surg., Dept. Navy. (ONR Contract No. N7-onr-450 11, NR 154 091.)

Received July 29, 1955

PSYCHOLOGICAL BULLETIN
Vol. 53, No. 4, 1956

CONCERNING KENDALL'S TAU, A NONPARAMETRIC CORRELATION COEFFICIENT

MAURICE S. SCHAEFFER
Los Angeles, California
AND EUGENE E. LEVITT[1]
Institute for Juvenile Research, Chicago

There have been several recent instances in the psychological literature (**1, 2, 6, 8, 9**) of the use of the statistic known as Kendall's tau (T), a nonparametric correlation coefficient. It is to be hoped that its use reflects a growing realization among psychologists of the inadequacy of the Pearson product-moment coefficient (r) in a number of circumstances. Some of these circumstances are:

1. When the variates to be correlated show sharp departures from normality. Although the distribution of sample r's from nonnormal but uncorrelated populations differs only slightly from the normal case (**4**), it may differ considerably when the true r is not zero, kurtosis rather than skewness being the more important factor (**3**).

2. When the variates to be correlated are unmeasureable according to an objective scale, as in the case of ratings or preferences of judges, or when precise measurement is impractical and the raw data must be sets of ranks. Under these circumstances, the evaluation and interpretation of r often requires assumptions which it would be imprudent to make.

3. When there is reason to believe that the regression of one variate on the other is nonlinear, r will tend to underestimate the degree of interdependence.

The use of a rank-correlation coefficient requires no assumptions regarding the form of the distributions of the variates and is thus admirably suited to the resolution of the difficulties posed by the first two circumstances. A rank coefficient also will not underestimate a relationship even when regression is nonlinear so long as the regression function is monotonic, which is usually the case in psychological research.

These considerations apply both to T and to the better-known rank correlation, Spearman's rho. This paper, however, will be concerned entirely with the former since it has a number of advantages over rho, and is rarely discussed in current statistical texts. The most important of these advantages is that the significance of a sample tau (τ) can be evaluated with certainty in terms of the normal probability integral for all but very small values of n. Furthermore, confidence limits for T can be determined from sample τ's. If the rank-order coefficient is regarded as merely a rough approximation of r, these considerations are not particularly important. When it is used as a test of an hypothesis for which it alone is appropriate, as is often the case, then these advantages, especially the former, become significant.

Tau can also be used for the computation of both partial and multiple correlation coefficients. However,

[1] The preparation of this paper was supported in part by Research Grants M-658C and MH-301 from the National Institutes of Health, U. S. Public Health Service. This work was done while both authors were at the Iowa Child Welfare Research Station.

neither of these measures is very useful at present as we shall indicate in a subsequent section.

Definition and Interpretation

Tau is defined as

$$\tau = \frac{S}{n(n-1)/2} = \frac{2S}{n(n-1)}, \qquad [1]$$

when n is the number of items ranked and $S=(P-Q)$, where P is the number of item pairs on the order of which both rankings agree, and Q is the number on which they disagree. Tau will vary from $+1.00$ when all possible pairings are ranking concordantly, to -1.00 when all pairings are ranked discordantly.

Consider the following rankings of the "ambiguity" of eight sentences made by two judges, with the rankings of Judge A arranged in the natural order:

Judge A	1	2	3	4	5	6	7	8
Judge B	5	1	6	7	2	3	8	4

The first sentence, i.e., the one ranked 1 by Judge A, has to its right in Judge B's rankings 3 larger ranks and 4 smaller ranks. We allot $+1$ for each of the larger ranks, and -1 for each of the smaller ranks. The second sentence in Judge A's ranking has to its right in Judge B's ranking 6 larger ranks and no smaller ranks. And so on through sentence 7. Alloting pluses and minuses in this fashion, we obtain: $+3$, -4; $+6$, -0; $+2$, -3; $+1$, -3; $+3$, -0; $+2$, -0; and $+0$, -1. P, the sum of the pluses is 17, and Q, the sum of the minuses is 11. S, which is $P-Q$, is thus 6. With $n=8$, we obtain according to [1] a τ of $6/28=.21$.

The interpretation of τ follows readily from [1] since $n(n-1)/2$ is the total number of item pairs with respect to which the rankings can be compared. A given value of T asserts that the statement "The order in which two items are ranked according to one variate (or judge) will be the order in which they are ranked by another variate (or judge)" will be correct $(100+100\text{T})/2$ per cent of the time, on the average.

When there are tied ranks, certain adjustments in the computational formula for τ must be made since the total number of possible item pairs will vary as a function of the ties. If there are ties in only one of the rankings, we arrange the untied ranking in the natural order and proceed to compute S as before except that the numbers in the second ranking, to the right of the items under consideration, which are the same as the rank of this item contribute nothing to the value of S. When both rankings contain ties, we arrange either one in the natural order and compute conventionally except that item pairs which are tied in the upper ranking also contribute nothing to the value of S.[2]

The major adjustment for tied ranks occurs in the denominator of [1] as might be expected. The general formula for τ from tied ranks containing the adjusted denominator is:

$$\tau = \frac{S}{\sqrt{[n(n-1)/2 - V]}\,\sqrt{[n(n-1)/2 - U]}}, \qquad [2]$$

[2] Smith's (**12**) description of the method for calculating τ when ties are present is in error since it neglects the effects of ties in the upper ranking on S. This oversight leads to markedly unreasonable τ's and distorts the sampling distribution by producing too many large absolute values of τ. For instance, in one of Smith's examples (**12**, p. 570), he obtains a corrected τ of $+1.00$ despite the fact that one judge perceived differences between items which were rated identically by the other. The correct procedure leads to a τ of .868, which expresses the high, though not perfect, degree of agreement which is present.

where

$$V=\tfrac{1}{2}\sum_{v} v(v-1),$$

and

$$U=\tfrac{1}{2}\sum_{u} u(u-1).$$

The computation of V and U will be illustrated in the following example.

In the two sets of ranks below, the upper ranking has been arranged in the natural order.

1	2.5	2.5	4	6	6	6	8
2	6	6	6	2	6	2	6

The first item, i.e., the one ranked 1 in the upper ranking, has five larger ranks to its right in the lower ranking, and none smaller. It is not tied with any other item in the upper ranking, so its contribution to S is $+5$. The second item has 2 smaller ranks to its right, and thus contributes -2. Although this item is tied in the upper ranking, the pairs which are tied are not involved in the contribution. Similarly for the third and fourth items. The fifth item has 2 larger ranks to its right, but one of these, the sixth item, is tied with the fifth in the upper ranking, and thus does not contribute. The net contribution of the fifth item is therefore only $+1$ instead of $+2$. A similar procedure for the sixth item leaves it with a net contribution of 0. The seventh item contributes $+1$. S, the net total, is $7-6=1$.

V and U are obtained in the following manner:' the upper ranking, from which V is computed, contains two sets of ties, one of extent 2 and one of extent 3. For the first set, $v=2$, and $v(v-1)=2(2-1)=2$. For the second, $v=3$, and $v(v-1)=3(3-1)=6$. The sum of the expressions $v(v-1)$ in the upper ranking is $(2+6)=8$, and $V=\tfrac{1}{2}(8)=4$. The lower ranking also contains two sets of ties, of extents 3 and 5. These will enter into the computation of U. For these ties, $u=3$ and 5 and $u(u-1)=6$ and 20. Hence $U=\tfrac{1}{2}(6+20)=13$.

Substituting the computed values of S, V and U in [2], we obtain $\tau=.05$.[3]

When τ is computed without adjusting the denominator for ties, it will always be numerically less than when the adjustments are made. The use of the uncorrected denominator is recommended by Kendall (7) when agreement with an objective ranking is being determined. In such a case, only the judge's ranking would contain ties, and these would theoretically indicate inability to discriminate the objective order, a failing for which the judge should properly be penalized. In general, however, the corrected formula should be used since rank correlations are usually computed when agreement rather than accuracy is the issue.

The procedure for adjusting for tied ranks can be generalized to include cases involving dichotomies. A dichotomy may be regarded as 2 sets of tied ranks of the extents of the number in each of the two categories, and the computational procedure need not differ from instances in which ties are less extensive. However, some labor can be avoided by the use of the following formulae:

$$\tau=\frac{S}{\sqrt{[n(n-1)/2-V]}\sqrt{pq}}, \qquad [3]$$

when one of the variates is a dichot-

[3] If the computation of τ when ties are present seems tedious, it should be pointed out that rho has no advantage in this respect. The proper computation of rho from tied ranks also involves corrections in both numerator and denominator, the latter being similar in form and effort to that required for τ. Unfortunately, most texts fail to mention, let alone describe, the corrections for rho, thereby creating the inaccurate belief that none are necessary.

omy consisting of p and $(n-p)=q$ members in the categories, or

$$\tau = \frac{S}{\sqrt{(pq)(xy)}}, \qquad [4]$$

when both variates are dichotomized into categories consisting of p and q members, and x and $(n-x)=y$ members. In this case, if we arrange the frequencies in a 2×2 table as for correlated proportions, S will be found to equal the difference between the products of the frequencies in the diagonal cells.

Tests of Significance[4]

The distribution of sample τ's for uncorrelated variables rapidly approaches normality and is satisfactorily approximated, when $n>10$, by the normal distribution with a mean of zero and a variance defined as

$$\sigma_\tau^2 = \frac{4n+10}{9n(n-1)}. \qquad [5]$$

When ties are present, the formula for the variance of τ becomes complicated.[5] If the number of ties is small, [5] may be used with only a slight error. Since the correction for ties will invariably reduce the variance, the use of the uncorrected formula will furnish a more conservative test of the null hypothesis.

Kendall (**7**) provides probability tables for evaluating the significance of an obtained S (rather than its τ) when $n \leqq 10$. Values of τ required for significance at the .10, .05, and .01 levels (or beyond, since τ can take only a limited number of values) for n's from 4 through 10 are shown in Table 1.

When ties are present in one of the rankings, Sillitto's tables (**11**) of the distribution of S for all possible numbers of pair and triplet ties for small n's may be used. When other types of ties are present, or when both rankings contain ties, the evaluation of τ is not feasible if n is 10 or less.

Correction for Continuity

When the significance of τ is evaluated using normal probability tables, it must be corrected for continuity, since S can not assume all values within the range $\pm\frac{1}{2}n(n-1)$. Since n is fixed, an increase in P is accompanied by a decrease in Q, and the

[4] In this paper *all* significance tests are attributable to Kendall (**7**) unless there is a specific indication to the contrary.

[5] The variance of τ when there are ties in both rankings is

$$\sigma_\tau^2 = \frac{2}{9n^2(n-1)^2}\Big\{n(n-1)(2n+5) - \sum_v v(v-1)(2v+5) - \sum_u u(u-1)(2u+5)\Big\} + \frac{4}{9n^3(n-1)^3(n-2)}\Big\{\sum_v v(v-1)(v-2)\Big\}\cdot\Big\{\sum_u u(u-1)(u-2)\Big\} + \frac{2}{n^3(n-1)^3}\Big\{\sum_v v(v-1)\Big\}\Big\{\sum_u u(u-1)\Big\}.$$

If only one ranking contains ties, this reduces to

$$\sigma_\tau^2 = \frac{2}{9n^2(n-1)^2}\Big\{n(n-1)(2n+5) - \sum_u u(u-1)(2u+5)\Big\}.$$

When one ranking is a dichotomy consisting of x and y members so that $(x+y)=n$, the variance is

$$\sigma_\tau^2 = \frac{4xy}{3n^3(n-1)^3}\Big\{n^3-n-\sum_u (u^3-u)\Big\}.$$

The variance when one ranking is a dichotomy and the other contains no ties is

$$\sigma_\tau^2 = \frac{4xy(n+1)}{3n^2(n-1)^2}.$$

When both rankings are dichotomies with x and y, and p and q members respectively, the variance becomes

$$\sigma_\tau^2 = \frac{4xypq}{n^2(n-1)^3}.$$

The above formulae are to be found in Kendall (**7**).

minimum change in S is thus 2. The appropriate correction for continuity is therefore to subtract 1 from the absolute value of S. This is equivalent to a deduction of $2/n(n-1)$ from the absolute value of τ, and the correction may be applied at either point.

This simple correction is appropriate when neither distribution contains ties, or when only one has ties. When one ranking consists entirely of ties of extent u, and the other ranking is a dichotomy, the correction for continuity consists of subtracting u from S, or $2u/n(n-1)$ from τ. If both variates are dichotomies, the deduction for continuity from S is $\frac{1}{2}n$ or $1/(n-1)$ from τ.

In instances where both rankings contain ties but are not dichotomies, there is no simple way of applying a correction. Whitfield's proposed correction (**13**) for the case in which one variate is a dichotomy and the other contains ties of varying extents might be used for the general case of ties in both rankings. Whitfield's method involves arranging the undichotomized ranking in the natural order and subtracting the extent of the ties involving the smallest and the greatest rank from twice the number of items ranked. This quantity is then divided by the number of intervals in the ranking. One-half of this quotient is the deduction from S for the correction. If τ is corrected instead of S, the deduction is the quotient divided by $n(n-1)$. The formal expression for this correction for S is

$$C_s = \frac{2n - v_1 - v_2}{2n_i}, \qquad [6]$$

where n is the number of items ranked, v_1 is the extent of the tie involving the smallest rank, v_2 is the extent of the tie involving the largest rank, and n_i is the number of intervals in the ranking. (If a ranking had no ties, $n_i = (n-1)$; in a dichotomy, $n_i = 1$.) In our illustrative problem (p. 340), $n = 8$, $v_1 = 1$, $v_2 = 1$, and $n_i = 4$. Accordingly, the deduction from S would be

$$\frac{(2 \times 8 - 1 - 1)}{2 \times 4} = 1.75.$$

The generalization of Whitfield's procedure to the general case of ties in both rankings is apparently not a simple matter, and it has not yet been accomplished. A suggestion would be to consider the ranking with the fewer intervals (and the most tied items) as a dichotomy, and to apply Whitfield's correction. This actually will provide an overcorrection for continuity and hence a safer test of the null.

Confidence Limits of T

It is often desirable to establish confidence limits for the parameter correlation when a significant sample coefficient has been obtained. For any value of a population **T**, the sampling distribution of τ tends rapidly toward normality (though not so rapidly as in the null case), provided that the absolute value of **T** is not too close to unity. The mean of the distribution is the population **T**, but the variance cannot be exactly determined unless something is known about the arrangement of ranks in the population, information which is almost always lacking. However, it can be shown that for any parameter **T**, the variance of τ cannot exceed the value

$$\text{maximum } \sigma_\tau^2 = \frac{2(1-\tau^2)}{n} \cdot \qquad [7]$$

Confidence limits of **T** can be set by substituting the value of the sample τ in [7]. An alternate method is to solve equation [8] with the roots providing the limits. The value of x is

the normal deviate corresponding to the desired probability level.

$$L_T = \frac{\tau + x\sqrt{\frac{2}{n}}\sqrt{1+\frac{2x^2}{n}-\tau^2}}{1+\frac{2x^2}{n}} \qquad [8]$$

Since the limits determined by means of [7] and [8] are based on a *maximum* variance, the probability is *at least*, but not precisely, $(1-P)$ that the true **T** lies within those limits. Unless n is fairly large, the magnitude of the limits will often be so great as to render them practically useless. Kendall (7) has developed an additional method which involves the estimation of a parameter representing the arrangement of ranks in the population from the obtained data. While this method frequently results in tremendous reductions in the extent of the confidence limits, it is too complicated and laborious for ordinary use.

Significance of a Difference Between τ's

Evaluating the significance of a difference between two independent τ's presents no special problems since such differences will be approximately normally distributed around a mean of zero in a test of the null hypothesis. The critical ratio which is conventionally used in such situations is applicable to τ. The standard error of the difference is, as usual, $\sqrt{\sigma_{\tau_1}{}^2+\sigma_{\tau_2}{}^2}$ where $\sigma_{\tau_i}{}^2$ is computed by [7].

If we wish to avoid using the sample τ as an estimate of **T** in computing the variances, we have recourse to a transformation called w, which is defined as $\sin^{-1}\tau$, in radians. Kendall (7) has shown that the sampling variance of w can be maximized at $2/n$, a value independent of the parameter w. The standard error of the difference between w_1 and w_2 can be maximized at

$$\sqrt{2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)},$$

an expression which does not require an estimation of population w's from the data.

The w transformation may also be used to set confidence limits for a **T**, though there is no reason to feel that this would be a desirable practice. Limits set in this manner, while differing slightly from those determined by [8], cannot be said to be more accurate, since it is not known whether the distribution of w is nearer normality than that of τ. Furthermore, the computations involved in converting from τ to w and back again may very well exceed those required in solving [8] to obtain the limits.

A Complete Computational Example

Consider the following set of rankings where the first has been arranged in the natural order:

1	2	3	4	5	6	7	8	9	10
6	8	10	9	7	5	2	4	1	3

Computing S, we obtain $+4$, -5; $+2$, -6; 0, -7; 0, -6; 0, -5; 0, -4; $+2$, -1; 0, -2; and $+1$, 0. The total for P is 9, the total for Q is 36, and $S=-27$. For the denominator, $\frac{1}{2}n(n-1)=\frac{1}{2}(10)(9)=45$. According to [1], $\tau=-27/45=-.60$. Entering Table 1 with an n of 10, we find that a τ of .60 is significant beyond the .05 level. The precise p value is .0166.

If we wish to use the normal approximation, we require the standard error of τ, and we must correct τ for continuity. From [5], we compute the variance of τ as .0617, and the standard error, .248. Applying the continuity correction at S, we recompute τ from [1] thus: $(-27+1)/45$

$=-.578$. Or, correcting τ itself; $-.60+2/(10)(9)=-.578$. (Since the correction is subtracted from the absolute value of S or τ, we add it to a negative statistic.) The critical ratio of τ is thus $-.578/.248=2.33$, which corresponds to a probability of .0198. Comparing this value with the probability obtained from Table 1, we see that the normal approximation is slightly in error when n is as small as 10, though it provides a somewhat more stringent null test.

To set the confidence limits of T at the .05 level and beyond, we solve [8] with $\tau=-.60$ and $x=1.96$. The roots of the quadratic are $-.93$ and $+.25$, which are the limits of T. The finding is hardly illuminating, though not unexpected. Any correlation based on only 10 instances is bound to be an uncertain estimate of the population value. If we had used [7] to compute the limits, we would obtain a maximum standard error of .358 and limits of $-.60\pm.70$ at the .05 level or beyond.

Partial Rank Correlation

A procedure for computing a partial τ when there are more than two rankings is described by Kendall (**7**). Suppose that we wish to determine the relationship between the rankings of Judges A and B with the ranking of Judge C held constant. Arrange the ranking of Judge C in the natural order, with those of Judges A and B beneath. There are $n(n-1)/2$ couplets in each ranking, i.e., items 1 and 2, 1 and 3, . . . 1 and n, 2 and 3, etc. In Judge C's ranking, the order of magnitude of each couplet is the same; the one to the right is the larger. We determine (a) the number of couplets on which both Judge A and Judge B agreed with Judge C as to order, (b) the number of couplets on which both disagreed with Judge C, (c) the number on which A agreed and B disagreed, and (d) the number on which B agreed and A disagreed.

These frequencies are now arranged in an ordinary 2×2 contingency table and the partial τ of the rankings of Judges A and B independent of that of Judge C is defined as

$$\tau_{AB\cdot C}=\frac{ab-cd}{\sqrt{(a+c)(a+d)(b+c)(b+d)}}. \quad [9]$$

It so happens that

$$\tau_{AB\cdot C}=\frac{\tau_{AB}-\tau_{AC}\tau_{BC}}{\sqrt{1-\tau_{AC}{}^2}\sqrt{1-\tau_{BC}{}^2}}, \quad [10]$$

an expression which is analogous to that for the product-moment partial correlation coefficient. It happens further that $\tau_{AB\cdot C}=\sqrt{x^2/n}$, which illustrates the relationship between partial τ and the phi coefficient.

Examples of the computation of partial τ using [9] can be found in Kendall (**7**) and Smith (**12**). The latter's example, though correct in form, contains arithmetic errors so that the computed partial is inaccurate. For

TABLE 1

Values of τ Required for Significance at the .10, .05 and .01 Levels and Beyond*

Level	n	4	5	6	7	8	9	10
.10		1.00	0.80	0.73	0.62	0.57	0.50	0.47
.05		—	1.00	0.87	0.71	0.64	0.56	0.51
.01		—	—	1.00	0.90	0.79	0.72	0.64

* Based on Kendall's (**7**) tables.

most purposes [10] will be the more useful computational method.

The use of partial τ when ties are present is questionable since [9] and [10] will give different results in such instances. This drawback, added to the fact that generally applicable tests of the significance of any partial τ are not yet available, limits the value of the statistic.[6]

An expression for a multiple τ has been developed by Moran (**10**), but the problems of the sampling distribution of multiple τ, although apparently simpler than those of partial τ, have also not yet been solved. The usefulness of multiple τ, like that of partial τ, is limited at the present time.

The Relationship Between τ and r

When ranked data can be assumed to be based on continuous, normal distributions and n is fairly large, an estimate of the parameter product-moment coefficient can be obtained by means of a transformation of τ. The formula for this transformation is

$$r = \sin \frac{\pi\tau}{2} \text{ (radians)}$$

$$= \sin 90\tau \text{ (degrees)}. \qquad [11]$$

The significance of the estimated r can be tested by simply testing the τ from which it was derived, using normal tables and a variance computed by [5].

In the nonnull case, the distribution of sample τ's will be approximately normal for large n's, with a mean of T and a maximum variance of

$$\text{maximum } \sigma_\tau^2 = \frac{5(1-\tau^2)}{9(1-\tau)} \cdot \qquad [12]$$

Confidence limits for T can be obtained using this variance, and corresponding limits for the transformed r are computed by translating the limits of T into those for r using [11].[7]

A comparison of the upper limit of the variance of τ by [12] when normality is assumed with its upper limit by [7] when no assumptions are made will show that the assumption of normality decreases the standard error of τ by approximately 50 per cent in the nonnull case. On the other hand, if τ is used to estimate r when the latter could be computed directly from the data, there will be a considerable loss of sensitivity since the standard error of the former is always greater than that of the latter. The ratio of the standard errors will vary from 1.2 when the variates are uncorrelated up to approximately 1.9 when the true r is .90.

The conversion formula for r from τ is justified only by the assumption of normality of distribution of the variates, and when n is fairly large. Otherwise, it would seem advisable to avoid estimating r from ranked data, and to limit the conclusions to statements concerning τ.

[6] Hoeffding (**5**) shows that when neither T_{AC} nor T_{BC} is unity, the distribution of $\sqrt{n}(\tau_{AB\cdot C} - T_{AB\cdot C})$ is approximately normal for large n's with a mean of zero and a variance given by an expression which he derives. Furthermore, when T_{AC} and T_{BC} are zero, the distribution of $\sqrt{n}(\tau_{AB\cdot C} - T_{AB\cdot C})$ is the same, in the limit, as that of $\sqrt{n}(\tau_{AB} - T_{AB})$.

[7] A standard error for r computed from τ can be derived using the conversion formula (**7**). Its upper limit is

$$\sqrt{\frac{1.37(1-r^2)(1-r^2)}{n-1}}.$$

The procedure for setting limits for r by converting limiting τ's into limiting r's is, however, preferable because of the greater symmetry of the distribution of τ.

REFERENCES

1. ADAMSON, R. E., & TAYLOR, D. W. Functional fixedness as related to elapsed time and to set. *J. exp. Psychol.*, 1954, **47**, 122–126
2. APPLEZWEIG, DEE G. Some determinants of behavioral rigidity. *J. abnorm. soc. Psychol.*, 1954, **49**, 224–228.
3. GAYEN, A. K. The frequency distribution of the product-moment correlation coefficient in random samples of any size drawn from non-normal universes. *Biometrika*, 1951, **38**, 219–247.
4. HALDANE, J. B. S. A note on non-normal correlation. *Biometrika*, 1949, **36**, 467–468.
5. HOEFFDING, W. A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.*, 1948, **19**, 292–325.
6. JONES, M. B. Authoritarianism and intolerance of fluctuation. *J. abnorm. soc. Psychol.*, 1955, **50**, 125–126..
7. KENDALL, M. G. *Rank correlation methods.* London: Griffin, 1948.
8. MALTZMAN, I., FOX, J., & MORRISETT, L. Some effects of manifest anxiety on mental set. *J. exp. Psychol.*, 1953, **46**, 50–54.
9. MONTGOMERY, K. C. Concerning the use of analysis of variance on latency data. *Amer. J. Psychol.*, 1953, **66**, 131–135.
10. MORAN, P. A. P. Partial and multiple rank correlation. *Biometrika*, 1951, **38**, 26–32.
11. SILLITTO, G. P. The distribution of Kendall's τ coefficient of rank correlation in rankings containing ties. *Biometrika*, 1947, **34**, 36–40.
12. SMITH, K. Distribution-free statistical methods and the concept of power efficiency. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences.* New York: Dryden, 1953.
13. WHITFIELD, J. W. Rank correlation between two variables, one of which is ranked, the other dichotomous. *Biometrika*, 1947, **34**, 292–296.

Received August 5, 1955

GEORGE BANTA COMPANY, INC., MENASHA, WISCONSIN